The Theory of Inbreeding

The
Theory of Inbreeding

By

Ronald A. Fisher, Sc.D., F.R.S.

Arthur Balfour Professor of Genetics, University of Cambridge
Foreign Associate United States National Academy of Sciences
and Foreign Honorary Member, American Academy of
Arts and Sciences; Foreign Member of the Swedish
Royal Academy of Science; formerly
Galton Professor, University
of London

CONTENTS

vii

I

INTRODUCTION

To the majority of practical breeders inbreeding appears as a rather mysterious danger. There is a reason for this, but there is no reason for neglecting the possibilities which the process opens out for the practical improvement of domestic plants and animals. Even if its importance could be ignored thirty years ago, at the time of the publication of East and Jones' pioneer book *Inbreeding and Outbreeding* (1), the resounding success of methods of maize improvement based on inbreeding, and the large contribution which this success has made to the prosperity of the United States, make it certain that a theoretical and practical study of the subject will form an essential part of future programmes either of genetic or of agricultural research.

The first important step towards the elucidation of the subject was taken by Darwin. Advancing from the notion that the innumerable adaptations of animals and plants were due to a selective process, whereby in each generation some kinds of individuals leave more descendants than others, he was struck by the many peculiarities of plants favouring cross-fertilisation rather than the self-fertilisation of which hermaphrodite plants are capable. Some plants, like so many animals, have come to have separate sexes. Others, while still hermaphrodite, bear the pollen and ovules on different inflorescences. In others again a self-sterility mechanism, genetically controlled by a series of multiple allelomorphs, prevents the acceptance of either type of pollen which the plant to be fertilised could itself produce. In yet other plants of many different orders,

there are two or three morphologically distinct forms adapted to cross-fertilisation, and incapable of legitimate self-fertilisation. The multiplicity of the means employed, and the similarity of the effect produced, provided for Darwin a strong argument that these diverse mechanisms had each been evolved in response to selection, by reason of some real advantage of the products of cross-fertilisation over those of self-fertilisation. The ancient and widespread incest taboos in man, and the production by many insects (and by woodlice) of unisexual families, strongly suggest that in animals with separate sexes, the mating of near relatives has been subject to a similar counter-selection, and that means have been found to avoid the frequent occurrence of such unions.

Guided by this inference, Darwin carried out a most important series of experiments to compare quantitatively the growth rates among the products of cross- and self-fertilisation in a number of plant species. The results were reported in his book (1876) on *The Effects of Cross- and Self-Fertilisation in the Vegetable Kingdom* (2). In all cases Darwin found, in accordance with previously expressed qualitative opinion, but for the first time with experimental accuracy, that the growth rate was generally higher in the offspring of crosses than in those of selfings. In this he justified the view that natural selection was competent to have brought about the numerous different adaptations favouring crossing. He did not, however, infer that the phenomenon he had demonstrated was due, as was commonly thought, to any permanent or cumulative injury to the stock. On the contrary he drew special attention to the facts that some individuals and their descendants could be found which showed no apparent signs of injury, and that the first cross-fertilised generation from apparently debilitated individuals

showed no signs of the inbreeding to which their parents individually had been subjected. The whole series of Darwin's experiments is consonant with the view that the injury done by self-fertilisation is immediate in affecting the offspring, but not permanent or cumulative, as it would be if it injured the stock.

The stages of the argument due to Darwin, are (i) mechanisms exist widely leading to the avoidance of inbreeding, and (ii) inbreeding does in fact have effects likely to incur a selective disadvantage. The question of how it has such effects was not answered by Darwin, and it is only since the rediscovery of Mendel's work that a rational theory has become possible. This theory depends essentially on the phenomenon of dominance, whereby the heterozygote, receiving different allelomorphic genes from its two parents, is not strictly intermediate between the two corresponding homozygotes, but resembles one of them, the *dominant*, often completely, while differing from the other, the *recessive*. Mendel enunciated no law of dominance, but introduced the terms dominant and recessive, to describe this phenomenon, which he encountered in all the seven factors studied by him in *Pisum*. The observation was of especial importance just because it does not follow as a logical consequence from the particulate theory of inheritance, embodied in his law of segregation, but is a fact additional to that law. Without it we could give no explanation of the effects of inbreeding, for the only effect which inbreeding can have, in accordance with the law of segregation, is to increase the frequency of the two kinds of homozygote equally at the expense of the frequency of the heterozygote. In the absence of dominance this can have no effect on the average, but can only increase the variability among the offspring of a self-fertilisation, or a union of close relatives. When, however, factors

showing dominance are present in the inbred material there will be an abrupt increase in the proportion of recessives, and these will differ phenotypically from the heterozygotes which they replace.

Genetic theory does, therefore, supply a rational basis for understanding the effects of inbreeding, provided we can answer the further question : Why should disadvantageous genes tend to be recessive ? In the chapter on the evolution of dominance in his book *The Genetical Theory of Natural Selection* (3), the author has given some account of the theory that disadvantageous genes are maintained in all species by rare mutations, the frequency of each (generally somewhat low) being such that the elimination by natural selection just balances fresh accretions by mutation ; that similar mutations, though rare events, have occurred regularly in the past, often for enormous periods ; that in this situation natural selection will constantly favour modifying factors tending to render each type of heterozygote more normal, or in other words to render the disadvantageous mutant more recessive ; and that such recessive genes, though each rare compared with its normal allelomorph, exist in great numbers in different parts of the germ plasm. This theory, although strenuously combated by Prof. Sewall Wright, appears now to be found generally acceptable. Since it was put forward, moreover, its only questionable premise, that the reaction of the heterozygote is readily modified by the selection of modifying factors, has been decisively verified by E. B. Ford with the magpie moth, *Abraxas grossulariata* (4), and by Fisher and Holt with the common house mouse, *Mus musculus* (5). There is no room to doubt that any other case chosen would show a correspondingly striking and rapid response.

If we add, then, to the primary genetical argument,

the view that deleterious genes, harmful to normal physiological efficiency, are constantly being supplied, at a low rate, by mutation, and that these have, by a slow evolutionary process, for the most part now become completely recessive to their normal allelomorphs, the effect of inbreeding will be seen to consist in uncovering, or exposing, some of the manifold defects of the germ-plasm, and we can understand both why cross-fertilisation mechanisms and instincts have come into existence to avoid this immediate evil ; and, on the other hand, why the germ plasm is completely uninjured, and why the deleterious effects produced should disappear entirely when inbreeding is followed by an outcross.

Practical breeders of farm animals are naturally deterred from a form of mating which is liable to produce animals undersized, unproductive, and liable to disease. Their reluctance is doubtless enhanced by a subconscious abhorrence of incest in their own species. Nevertheless, when an inbred line is formed from élite stock, it can contain no genes, however inferior it may appear, which were not present in its admired progenitors, nor can it hand any others on to its descendants. At the expense of some loss of appearance, and immediate utility, and with the real inconvenience of lower fertility, which may make the maintenance of such stocks difficult, the germ plasm may have been purified of many unnecessary defects, and the great boon of reliability of breeding performance gained. All that is learnt of the breeding potentialities and environmental reactions of such stocks is knowledge of continuing value, permanently applicable to practical purposes so long as the inbred stock is maintained. Slow as the production of such closely inbred stock undoubtedly is, the forty-nine years which have elapsed since the rediscovery of Mendel's work would have sufficed to supply the present generation of animal breeders with

material of the utmost value, had the adoption of inbreeding programmes not been delayed by prejudice and inertia, reinforced perhaps by superstitious fear. The example of their success in plant breeding is, however, now so overwhelming that serious steps will certainly be increasingly taken to make similar improvements in the breeding of livestock.

REFERENCES

(1) E. EAST and D. F. JONES (1919). Inbreeding and outbreeding, their genetic and sociological significance. Washington Square Press, Philadelphia.

(2) C. DARWIN (1876). The effects of cross- and self-fertilisation in the vegetable kingdom. John Murray, London.

(3) R. A. FISHER (1930). The genetical theory of natural selection. Clarendon Press, Oxford.

(4) E. B. FORD (1940). Genetic research in the *Lepidoptera*. Galton lecture 1939 to the Seventh International Congress of Genetics. Ann. Eugen., x. 3, 227-252.

(5) R. A. FISHER and S. B. HOLT (1944). The experimental modification of dominance in Danforth's short-tailed mutant mice. Ann. Eugen., xii. 2, 102-120.

SEGREGATING INBRED LINES

1. Formal Structure

FOR genetical purposes it is important that a line can be inbred while still maintaining segregation in one or more manifest factors. In each generation matings are selected so that a heterozygote is mated to a recessive

$$Aa \times aa.$$

Such a mating will reproduce the genotypes of the parents in equal numbers. The progress of inbreeding depends only on the mating of close relatives, so that if a male and female of different genotypes are selected, the segregation will be maintained and the inbreeding continued. With animals such as mice bearing several at a birth, such selection is easy. Half the possible matings in the long run will be between mice of complementary types, and such matings may be spoken of as eligible for the continuation of the line.

The possibility of maintaining such a line with a number of different factors segregating simultaneously depends on the prospect of obtaining eligible matings. We shall first give a calculation of the number of eligible matings to be expected in a line producing N genotypes of each sex in equal numbers.

2. The Number of Eligible Matings Expected

Of s mice (or whatever the organism may be) bred let t be males, and let these comprise g of the N genotypes available for each sex ; then, supposing each male can be used with any number of females, each of the $s-t$ females has a probability g/N of finding a

suitable mate, and the expected number of eligible matings is

$$\frac{g}{N}(s-t).$$

For any given value t, the probability of any value g is calculable from the so-called differences of the powers of zero, and is

$$\frac{1}{N^t} \cdot \frac{N!}{g!(N-g)!} \Delta^g u_0,$$

where $u_x = x^t$; then the expectation of g/N is

$$\frac{1}{N^t} \sum_{g=1}^{N} \frac{(N-1)!}{(g-1)!(N-g)!} \Delta^g u_0 = \frac{1}{N^t} \Delta(N-1)^t = 1 - \left(1-\frac{1}{N}\right)^t.$$

Less formally this may be seen from the fact that the chance of any chosen male genotype being unrepresented is

$$\left(1-\frac{1}{N}\right)^t.$$

But for any given value of s the probability of each possible value t is given by the binomial expansion of $(\frac{1}{2}+\frac{1}{2})^s$, and is

$$\frac{1}{2^s} \cdot \frac{s!}{t!(s-t)!} ;$$

hence the expectation of the number of eligible matings is

$$\frac{1}{2^s} \sum_{t=1}^{s} \frac{s!}{t!(s-t-1)!} \left\{1-\left(1-\frac{1}{N}\right)^t\right\}$$

$$= \frac{s}{2}\left\{1-\frac{1}{2^{s-1}}\left(1+1-\frac{1}{N}\right)^{s-1}\right\}$$

$$= \frac{s}{2}\left\{1-\left(1-\frac{1}{2N}\right)^{s-1}\right\}$$

Table 1 gives the values of these expectations for N equal to 8, 16, 32 corresponding with 3, 4 and 5 independently segregating factors.

TABLE I

Values of $\frac{1}{2}s\left\{1 - \left(1 - \dfrac{1}{2N}\right)^{s-1}\right\}$,

$s =$	$N = 8.$	$N = 16.$	$N = 32.$	$1 = s$
1	·0	·0	·0	1
2	·0625	·03125	·015625	2
3	·1816466	·092285	·046509	3
4	·3520508	·181702	·092293	4
5	·5688095	·298154	·152626	5
6	·8274107	·440354	·227164	6
7	1·1237305	·607067	·315572	7
8	1·4539969	·797110	·417518	8
9	1·8147624	1·009351	·532681	9
10	2·2028775	1·242704	·660745	10
11	2·6154674	1·496131	·801400	11
12	3·0499098	1·768639	·954034	12
13	3·5038157	2·059274	1·119282	13
14	3·9751606	2·367128	1·295921	14
15	...	2·691327	1·483979	15
16	...	3·031037	1·683178	16
17	...	3·385462	1·893246	17
18	...	3·753838	2·113916	18
19	...	...	2·344929	19
20	...	...	2·586028	20
21	...	...	2·836965	21
22	...	...	3·097495	22
23	...	...	3·367380	23
24	...	...	3·646384	24
25	...	...	3·934281	25

3. Linked Factors

If two of the factors used are linked, and are both heterozygous in the same parent, half the genotypes will be due to recombination. Both or neither of the genotypes of an eligible mating must be in this group. We may now calculate the expectation of eligible matings of these two classes.

If s' is the number of animals bred in the recombination classes, and n is the number of recombination genotypes in both sexes together, the expected number of suitable matings is

$$\tfrac{1}{2}s'\left\{1 - \left(1 - \dfrac{1}{n}\right)^{s'-1}\right\};$$

B

but out of s animals bred the probability of s' recombinations is

$$\frac{s!}{s'!(s-s')!}\, p^{s'}q^{s-s'},$$

if the fraction p are expected to be recombinants, and q non-recombinants. Hence the expected number of suitable matings of the recombination classes from s animals bred is

$$\tfrac{1}{2}\sum_{s'=1}^{s}\frac{s!}{(s'-1)!(s-s')!}\, p^{s'}q^{s-s'}\left\{1-\left(1-\frac{1}{n}\right)^{s'-1}\right\}$$
$$=\tfrac{1}{2}ps\left\{1-\left[p\left(1-\frac{1}{n}\right)+q\right]^{s-1}\right\}$$
$$=\tfrac{1}{2}ps\left\{1-\left(1-\frac{p}{n}\right)^{s-1}\right\}.$$

Equally the expected number of suitable matings of non-recombinational genotypes is

$$\tfrac{1}{2}qs\left\{1-\left(1-\frac{q}{n}\right)^{s-1}\right\}.$$

If from a mating in coupling a non-recombinational pair is selected the new mating will also manifest linkage in the coupling phase. If a recombinational pair is selected, each will be heterozygous for one of the linked factors, and from them all genotypes will appear in equal numbers. In half the eligible matings from among these, this condition will be continued. In the other half the matings will exhibit linkage, this time in repulsion. From repulsion progenies we can make up coupling matings by choosing recombinations, or non-linkage matings by choosing old combinations. Thus :—

$$N\left\langle\begin{array}{l}N\\R\text{---}N\\ \quad C\text{---}C\\ \qquad N\end{array}\right.$$
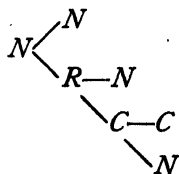
where the matings made by continuing a linkage mating using old combinations are shown in the same line. Note that repulsion matings can only follow matings of the non-linkage type. The number of non-linkage matings must be at least equal to the number in repulsion in any recurring system.

4. The Probability of at Least One Suitable Mating

The expected number of eligible matings gives useful guidance as to the numbers likely to be needed in continuing a segregating inbred line. It may, however, be asked further how often we shall fail to obtain any eligible mating after breeding a given number of young.

If s objects are distributed among an even number $2N$ of equal cells, to each of which a second corresponds reciprocally; the chance of any chosen cell being unoccupied is

$$\left(1 - \frac{1}{2N}\right)^s,$$

and the chance of any chosen set of t cells being unoccupied is

$$\left(1 - \frac{t}{2N}\right)^s.$$

Hence the chance that one or the other of any pair of cells is unoccupied is

$$p_1 = 2\left(1 - \frac{1}{2N}\right)^s - \left(1 - \frac{2}{2N}\right)^s.$$

Now the chance that any t of $2r$ chosen cells shall be unoccupied, and the remaining $2r - t$ be occupied is

$$\left(1 - \frac{t}{2N}\right)^s - (2r - t)\left(1 - \frac{t+1}{2N}\right)^s + \frac{(2r-t)(2r-t-1)}{1 \cdot 2} \times \left(1 - \frac{t+2}{2N}\right)^s$$

$$- \ldots (-)^{2r-t}\left(1 - \frac{2r}{2N}\right)^s.$$

Hence the chance that any two chosen pairs are each not both occupied is

$$p_2 = 4\left(1 - \frac{2}{2N}\right)^s - 4\left(1 - \frac{3}{2N}\right)^s + \left(1 - \frac{4}{2N}\right)^s,$$

and for three chosen pairs

$$p_3 = 8\left(1 - \frac{3}{2N}\right)^s - 12\left(1 - \frac{4}{2N}\right)^s + 6\left(1 - \frac{5}{2N}\right)^s - \left(1 - \frac{6}{2N}\right)^s$$
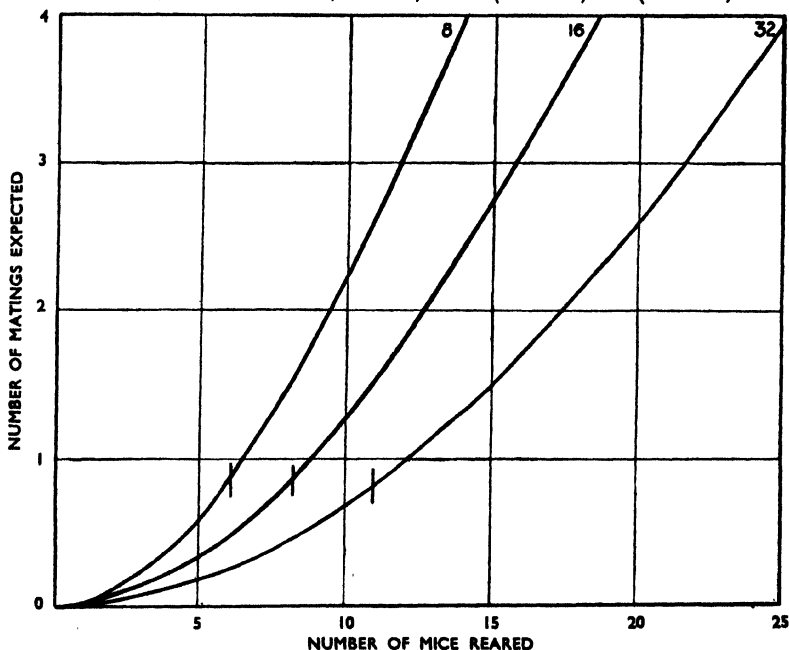
FIG. 1.—Number of appropriate matings expected in lines segregating 8, 16 and 32 genotypes of each sex plotted against the number of mice reared from any mating.   The vertical lines indicate the average number of mice needed to obtain one appropriate mating.

or, if $u_x = 2^x x^s,$ $p_r = \dfrac{1}{(2N)^s \cdot 2^{2N-2r}} \Delta^r u_{2N-2r},$

representing the chance that any chosen r of the N pairs of reciprocal eligible matings shall not be available.

The probability of no eligible pair is, therefore

$$p_N(s) = \frac{1}{(2N)^s} \Delta^N u_0 = \frac{1}{(2N)^s} \sum_{r=0}^{N} \frac{N!}{r!(N-r)!} (-)^r 2^{N-r} (N-r)^s.$$

These probabilities are somewhat laborious to calculate. For small values of s they may be expressed in a few terms only, *e.g.*

s	
2	$1 - \dfrac{1}{2N}$
3	$1 - \dfrac{3}{2N} + \dfrac{3}{4N^2}$
4	$1 - \dfrac{6}{2N} + \dfrac{15}{4N^2} - \dfrac{13}{8N^3}$
5	$1 - \dfrac{10}{2N} + \dfrac{45}{4N^2} - \dfrac{95}{8N^3} + \dfrac{75}{16N^4}.$

For high values of s, larger than N, when the probability is approaching zero, if no eligible mating has yet appeared, nearly half of the $2N$ genotype classes will have been filled, their complementary classes being still empty ; consequently the probability that the next animal classified will complete a pair must rise towards a half, and the probability of failing to make a mating will decrease nearly to half its previous value for each additional animal.

For example, with $N = 4$, the probabilities are :—

TABLE 2

Number Bred (s).	Probability of No Eligible Mating $p_4(s)$.	Probability of Completing the First Eligible Mating with the s-th Offspring $p_4(s-1) - p_4(s)$.
2	·875	·125
3	·672	·203
4	·459	·213
5	·286	·173
6	·161	·124
7	·093	·068
8	·050	·043
9	·027	·024
10	·014	·013
11	·007	·007
12	·004	·003

Both series, $p_N(s)$ and $p_N(s-1)-p_N(s)$, alike tend to zero in equality with 2^{N-s}.

If the young are produced one at a time, the time consumed in waiting for an eligible mating would be proportional to the average number that would require to be bred, that is to the mean of the frequency distribution shown in the third column of Table 1.

5. The Mean Number of Animals which Need be Bred

The mean and indeed the other cumulants of this distribution may be found from the expression already given for the probability ; for, since the probability of completing the first eligible mating with the s-th offspring is

$$p(s-1)-p(s),$$

the mean of the distribution must be

$$\sum_{s=1}^{\infty} s(p(s-1)-p(s)) = \sum_{s=0}^{\infty} p(s),$$

and the factorial moments about zero may be found similarly, from the relations

$$\sum_{s=1}^{\infty} \frac{s(s-1)}{2}(p(s-1)--p(s)) = \sum_{0}^{\infty} sp(s)$$

$$\sum_{s=1}^{\infty} \frac{s(s-1)(s-2)}{3!}(p(s-1)-p(s)) = \sum_{0}^{\infty} \frac{s(s-1)}{2!}p(s).$$

Now, with $$p(s) = \sum_{r=0}^{N} \frac{N!}{r!(N-r)!}\cdot(-)^r 2^{N-r}\cdot\left(\frac{N-r}{2N}\right)^s,$$

we find $$\sum_{s=0}^{\infty} p(s) = \sum_{r=0}^{N} \frac{N!}{r!(N-r)!}\cdot(-)^r 2^{N-r}\cdot\frac{2N}{N+r},$$

$$\sum_{s=0}^{\infty} sp(s) = \sum_{r=0}^{N} \frac{N!}{r!(N-r)!}\cdot(-)^r 2^{N-r}\cdot\frac{2N(N-r)}{(N+r)^2},$$

$$\sum_{s=0}^{\infty} \frac{s(s-1)}{2}p(s) = \sum_{r=0}^{N} \frac{N!}{r!(N-r)!}\cdot(-)^r 2^{N-r}\cdot\frac{2N(N-r)^2}{(N+r)^3},$$

and so on, in general.

These may all be determined arithmetically for any chosen N by summation of the terms derived from the expansion $(2-1)^N$. In the case of the mean it has been shown by K. Williams that a simple algebraic form can be found.

For any given function of r, which we may write v_r, we may use the equivalence

$$\sum_{r=0}^{N} \frac{N!}{r!(N-r)!} (-)^r 2^{N-r} v_r = (1-\Delta)^N v_0,$$

but, in the case of the function

$$v_r = \frac{2N}{N+r},$$

we have the simple relation

$$\Delta^r v_0 = 2N(-)^r \cdot \frac{r!(N-1)!}{(N+r)!} = (-)^r 2 \cdot \frac{r!N!}{(N+r)!};$$

hence

$$(1-\Delta)^N v_0 = 2 \sum_{r=0}^{N} \frac{N!}{r!(N-r)!} \cdot \frac{r!N!}{(N+r)!} = 2 \sum_{r=0}^{N} \frac{(N!)^2}{(N-r)!(N+r)!}$$

$$= 1 + \sum_{s=0}^{2N} \frac{(N!)^2}{s!(2N-s)!},$$

putting $s = N-r$, and separating half the term for $r = 0$,

$$= 1 + \frac{2^{2N}(N!)^2}{(2N)!},$$

which is the average ordinal value of the first eligible mating as the young are examined one by one. If the young come in litters of average size l, we must add something like $\frac{1}{2}(l-1)$ to find the average number of young bred when the first eligible mating becomes available. With pigs or mice we might then add 3 or 4

to the numbers given in Table 3, calculated from the formulæ above. The average number of litters required should not then exceed two for these species, even if five factors are kept segregating.

TABLE 3

Average Number of Young Needed to Obtain One Eligible Mating, with the Sampling Variance and Standard Deviation

N.	Mean.	Variance.	Standard Deviation.
4	4·65714	4·42925	2·06137
8	6·09215	7·59484	2·75587
16	8·14541	14·40452	3·79533
32	11·06575	28·10085	5·30102

For large values of N, beyond those tabulated, a satisfactory approximation is afforded by the distribution

$$p = e^{-(s-1)^2/4N}.$$

The value

$$s = 1 + \sqrt{12N}$$

will reduce p to about 5 per cent., so that this number will seldom need to be exceeded. The mean is

$$1 + \sqrt{\pi N},$$

and the variance

$$(4-\pi)N.$$

These approximate values should be compared with the exact values for $N = 32$.

6. Uses of Segregating Inbred Lines

Segregating inbred lines are needed to determine the actual effects of manifest genetic differences, for in any individual organism the manifestation must depend on the other genes present. This requires not only

that the genotypes to be compared shall be alike in respect of other manifest factors affecting the same organ or organ system, but that they shall be really comparable in all genetic respects, so that we shall not ascribe to any one factor under examination the joint consequences of it together with numerous unrecognisable factors, in which different stocks normally differ. Such comparability can only be assured within an inbred line. It is, of course, of especial importance in studies of dominance, for no statement of dominance is reliable without a valid comparison of the heterozygote with both of the corresponding homozygotes. It is also of importance whenever genetic differences are to be demonstrated either to students, or in a permanent record such as illustration of books, or museum specimens. For such purposes lines segregating in several factors are particularly desirable, for with these not only single contrasts, but the interactions of two or more can be exhibited.

The preservation of living genetic material is also best carried out in segregating inbred lines. Genetic progress with most species in the past has consisted in the recognition of the effects of some few, or many, genetic factors. Future work requires that these shall be preserved, and in a form available for immediate use. If this is not done, the value of past work is thrown away. To preserve a rare gene in an isolated stock enables it to be recovered for interbreeding or comparison with another, or the same, gene discovered later, but only at the expense of several generations of preliminary crossing to introduce the normal allelomorph, and to free the cross from unwanted factors. What is needed in practice is the gene and one at least of its allelomorphs, usually its normal allelomorph, segregating together on a uniform genetic background, and this is supplied by a segregating inbred line. If the gene is segregating as a

dominant, we are at once in a position to introduce the new gene in a parallel line ; if as a recessive, we should first have to make up homozygotes of the normal gene in the chosen line before introducing the new gene. The time consumed in preparing genetic material for making an adequate comparison is often considerable, and it is a great advantage to possess stocks of prepared material available for future use.

Inbred lines segregating in several factors ensure a steady supply of data on possible linkages. Close linkage is usually discovered quickly ; loose linkage is liable to be overlooked, and requires a large number of progeny for its demonstration. Partial linkage with sex is to be expected in many more organisms than have yet shown it (*Lebistes* and Man), and until recently no sex linkage had yet appeared in the large amount of study devoted to the rodents. When loose linkage is in view the accumulation of negative evidence for factors in different linkage groups is as important as the discovery of new linkages, for such data are often not reported. A line segregating in five factors, supposedly unlinked and autosomal, will, as it is continued, supply evidence relevant to ten possible autosomal linkages and to five possible sex linkages. Every mating will give information on some of these, either in coupling or repulsion. It is to be noted that all eligible matings are not equally informative. Thus a five-fold dominant male mated to a five-fold recessive female will be a double backcross for all fifteen items, while a doubly dominant male mated to a triply dominant female will supply evidence on only six. When a choice has to be made between different eligible matings this circumstance may be borne in mind, although until the material is already highly inbred, it would be unwise to delay the inbreeding programme in order to in-crease the information as to linkage. For five factors

the possible mating types may be classed as follows :—

TABLE 4

		Autosomal Linkage.	Sex Linkage.	Number of Mating Types.
abcde	♀×*ABCDE* ♂	10	5	1
Abcde	♀×*aBCDE* ♂	6	4	5
ABcde	♀×*abCDE* ♂	4	3	10
ABCde	♀×*abcDE* ♂	4	2	10
ABCDe	♀×*abcdE* ♂	6	1	5
ABCDE	♀×*abcde* ♂	10	0	1

On the average of all matings, each contributes to five autosomal and two-and-a-half sex linkages. As each mating is completed its contribution should be carried through to the summary of matings in coupling and repulsion for each linkage being tested.

Inbred lines are conveniently carried on in a series of parallels. How many must depend on the probability of needing a rapid expansion of the stock. For security a reasonable number is three, if each pair is discarded as soon as its offspring have bred. The three lineages are then continued in parallel, and on the failure of any one it is replaced by a pair from one of the others. Crossing between different lineages should not ordinarily be practised. If a large number of uniform young are required from an outcross with another line, steps should be taken in time to set aside the necessary parents. Lines segregating in few factors or none will build up such a supply most quickly, since it is only the homozygotes which when outcrossed will give a genetically uniform progeny. With experience, however, others carrying one or more recessives will certainly be found to introduce no appreciable variability for the purposes, such as biological assay, for which they may be required.

Sets of lines may be made up surveying comprehensively the linkages of a number of apparently independent factors, and, at the same time, though

less comprehensively, their phenotypic interactions. With mice, for example, there are 20 pairs of chromosomes, including the sex pair. We may also find 21 factors supposedly all unlinked with each other and with sex. If each is used in 5 lines, with all 4 associates different for every line, it will be used in one line with each of the 20 other factors. In 21 lines, then, the 210 autosomal linkages will all be tested once, and the 21 sex linkages each 5 times. Since there are only 20 chromosomes, 2 of the linkages at least must be real. The combinatorial problem of allocating the factors to the lines has been solved. It is equivalent to one of the problems of incomplete blocks arising in the design of controlled experiments. This solution is based on the orthogonal 4×4 square. The most ready method of arriving at a solution in practice is to arrange the factors in any order in a circle, and counting from any one of the 21 possible starting points, to choose numbers 1, 4, 5, 10, 12 as given in Fisher and Yates' *Statistical Tables*, where other aspects of incomplete blocks and orthogonal squares are also discussed. It may, however, easily be verified that this method will give 21 selections of the kind required, and a wider knowledge is needed only for adapting the method to other organisms. When, as in this case, a cyclic solution exists, a stencil may be cut so as to expose the chosen sets of 5 when the factors to be used are written down in a circle with 21 sectors, or in a column, with the first 11 repeated
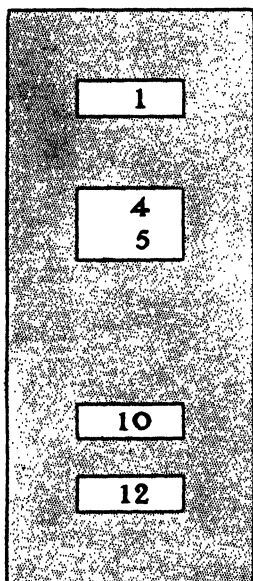
at the end of the list. This process is widely applicable, and deserves to be more used in experimental work.

With factors having more than two allelomorphs, any two may be chosen, consonantly with the efficient classification of the other factors present. Equally, if two or more factors are known to be linked, either may be chosen in any given line. It is thus a real advantage that each linkage group should appear in a number of lines. For example, in mice the albino factor has the four allelomorphs, C for full colour, c^{ch} chinchilla, c^e extreme dilution and c albino. It is also linked with sh_1 shaker, and with p pink-eyed dilution. The five lines in which this group is to appear might contain

$$Cc^{ch} \quad \times \quad c^{ch}c^{ch}$$
$$c^{ch}c^e \quad \times \quad c^{ch}c^{ch}$$
$$c^{ch}c \quad \times \quad c^{ch}c^{ch}$$
$$Pp \quad \times \quad pp$$
$$Sh_1sh_1 \times sh_1sh_1,$$

so including all eight genes; it will be noticed that the albino and the extreme dilution homozygotes are avoided as incompatible with the classification of other colour factors. This is done by manifesting c^e and c by their dominant effects in relation to chinchilla. Extremely dilute pigmentation which would mask other diluting factors may thus be avoided. If " intense chinchilla " really exists and should chance to be rediscovered, it could be accommodated by substitution for chinchilla in one of the first three lines.

When a linkage is observed, five of the lines, in which one of the linked factors were used, must be assigned to some new factor awaiting test, the five in which the

other factor was used, including that in which both occurred, will be divided as seems most convenient between the two factors now known to be linked. Sometimes both may be used together. In this way a new linkage will be created, unless more than one previously existing remains undiscovered.

III

PROGRESS TOWARDS HOMOZYGOSITY

7. The Mating System

In the previous chapter we have assumed the use of sib-matings, whole brothers and sisters, as the most convenient form of inbreeding for animals having many at a birth. It is not, however, the only possible form, and does not possess all the advantages of other methods in special circumstances. Some of these will be described in the following chapter (IV). It is, however, first necessary to develop a sufficiently comprehensive approach to the mathematical problems which arise when any mating system is adopted, and to those which arise when it is changed or varied. For this purpose we may first consider in some detail the genetic consequences of repeated sib-matings.

8. The Generation Matrix

Confining attention to a single locus, *individuals* may be of only two classes, homozygotes and heterozygotes, but the number of different types of *mating* to be distinguished is seen on consideration to be seven. These are :—

(i) Like homozygotes, represented by *aa* × *aa*, yielding only the genotype *aa*, and so leading to only one type of mating, namely the parental type (i).

(ii) Heterozygote mated to a homozygote having one gene in common with it, as *aa* × *ab* ; this will yield the two parental types of offspring *aa* and *ab* in equal numbers, and among these

three types of mating may be made, namely
$aa \times aa$ or (i) with frequency $\frac{1}{4}$, $aa \times ab$ or (ii)
with frequency $\frac{1}{2}$, and finally $ab \times ab$ or (iii)
with frequency $\frac{1}{4}$.

(iii) Two like heterozygotes $ab \times ab$ yielding $\frac{1}{4}$ aa,
$\frac{1}{2}$ ab and $\frac{1}{4}$ bb. If these are mated at random
we shall find $\frac{1}{8}$ of type (i) namely $\frac{1}{16}$ $aa \times aa$
and $\frac{1}{16}$ $bb \times bb$, next $\frac{1}{2}$ will be of type (ii) namely
$\frac{1}{4}$ $aa \times ab$ and $\frac{1}{4}$ $ab \times bb$, then, $\frac{1}{4}$ will be of
type (iii), and, finally $\frac{1}{8}$ will be $aa \times bb$, type (iv).

(iv) Unlike homozygotes $aa \times bb$; these will yield
only ab and give rise only to matings between
like heterozygotes (iii).

(v) Two heterozygotes having one gene in common
$ab \times ac$. These will yield $\frac{1}{4}$ aa, $\frac{1}{4}$ ab, $\frac{1}{4}$ ac, $\frac{1}{4}$ bc.
If these are mated at random, we have $\frac{1}{16}$
type (i), $\frac{1}{4}$ type (ii), $\frac{3}{16}$ type (iii), $\frac{3}{8}$ type (v)
and $\frac{1}{8}$ $aa \times bc$ type (vi).

(vi) Heterozygote with homozygote having no gene
in common with it, $aa \times bc$. This will yield
ab and ac in equal numbers. The matings
between these are $\frac{1}{2}$ of type (iii) and $\frac{1}{2}$ of
type (v).

(vii) Two heterozygotes having nothing in common,
$ab \times cd$, yielding four genotypes all heterozygous
in equal numbers, ac, ad, bc, bd. On inter-
breeding these at random we have $\frac{1}{4}$ type (iii),
$\frac{1}{2}$ type (v) and $\frac{1}{4}$ type (vii).

The consequences of each possible type of mating
may thus be followed out, and we find that each type
except the first leads to one or more of the other types
in a fixed proportion of cases. The whole of the genetic
consequences of inbreeding may be embodied in a table
showing the frequency with which each type leads to
each. If we knew the probabilities of each of the

seven types in any one generation we could infer the probabilities of each in the next, and, by repeating the process, in all subsequent generations in which the same breeding system was employed.

If then t, u, v, w, x, y and z stand for the frequencies of the different mating types in any generation, we may use a suffix o for the initial condition, 1, 2, ..., n, for subsequent generations, and can write down 7 equations expressing t_1, u_1, v_1, w_1, x_1, y_1, and z_1 in terms of t_0, u_0, v_0, w_0, x_0, y_0, and z_0. We shall adopt a tabular form for expressing the genetic facts, which may easily be read as such a system of equations.

TABLE 5

One Generation of Sib-mating

	$\dfrac{t_0}{1}$	$\dfrac{u_0}{4}$	$\dfrac{v_0}{8}$	$\dfrac{w_0}{1}$	$\dfrac{x_0}{16}$	$\dfrac{y_0}{2}$	$\dfrac{z_0}{4}$
t_1	1	1	1	.	1	.	.
u_1	.	2	4	.	4	.	.
v_1	.	1	2	1	3	1	1
w_1	.	.	1	.	.	.	.
x_1	.	.	.	.	6	1	2
y_1	.	.	.	.	2	.	.
z_1	.	.	.	.	.	.	1

Each column of the table represents one of the original mating types, and shows how frequently the next mating derived from it will be of any of the possible kinds. The total of the entries in any column appears as a divisor of the symbol standing at its head, so that the rows of the table may be read as equations ; *e.g.*

$$t_1 = t_0 + \tfrac{1}{4}u_0 + \tfrac{1}{8}v_0 + \tfrac{1}{16}x_0$$

representing the fact that t_1 will be supplied by t_0, u_0,

v_0, and x_0, but not by other initial mating types, and that all the matings derived from t_0, $\frac{1}{4}$ of those from u_0, $\frac{1}{8}$ from v_0, and $\frac{1}{16}$ from x_0 will be matings between like homozygotes.

Since t continually increases, while the frequencies of all the other types must diminish, and since the gain in t must exactly balance the losses in the others, we may in practice ignore the first row and the first column of the table, leaving a 6×6 table for the other mating types. We shall still, of course, take account of the divisors at the column heads. Apart from the inconvenience in writing and printing, these divisors could be brought into the body of the table, which would then be a 6×6 matrix of the coefficients of $t_0 \ldots z_0$ in the expressions for $t_1 \ldots z_1$. We shall refer to this matrix, or 6×6 table, as the generation matrix, for it shows the exact course of the changes to be expected in the passage of one generation.

It will be noticed that the table of the generation matrix has been divided horizontally and vertically into four sections corresponding with matings involving 1, 2, 3 and 4 different genes respectively. Since it is impossible for this number to increase in the course of a generation, the occupied portion of the table will consist of a chain of blocks containing the diagonal, representing cases in which the gene number is unchanged, and of the blocks standing above these, representing the cases in which the gene number is diminished. The blocks below, representing an increase in gene number are necessarily empty. This subdivision is a convenience for practical solution, but the properties with which we shall be concerned are in general theory to be regarded as properties of the 6×6 matrix as a whole.

Seeing that the generation matrix supplies the means of expressing the frequencies $u_1 \ldots z_1$ in terms of

u_0 . . . . z_0, and a similar set of equations for t_2 . . . . z_2 in terms of t_1 . . . . z_1, by substituting in the latter the expressions given in the former, we could express t_2 . . . . z_2 as linear functions of t_0 . . . . z_0, and thus have a matrix showing the results of two generations' inbreeding. Mathematically this is known as the square of the generation matrix, and if the process were carried on to n generations, so as to express u_n . . . . z_n in terms of u_0 . . . . z_0, the table obtained would be the n-th power of the original matrix. The process we shall discuss is therefore equivalent to that of finding the n-th power of any given matrix.

9. The Latent Roots

Turning now to the particular problem presented by sib-mating, we may notice that the equation for z_1 is of the simple form

$$z_1 = \tfrac{1}{4}z_0,$$

which admits of the easy generalisation

$$z_n = (\tfrac{1}{4})^n z_0.$$

The frequency of matings of this type, which we may now conveniently call type z, $ab \times cd$, will therefore diminish rather rapidly. After five generations, since $(\tfrac{1}{4})^5$ is less than a thousandth, certainly not more than one in a thousand inbred matings will be of this type. Expressing the matter otherwise we may say that not more than one thousandth part of the germ plasm will be in this condition after five generations, and after ten generations not more than one millionth.

The expressions for the other frequencies are more complex. It is possible, however, to imagine that the frequencies of these might be so adjusted that their relative frequencies remained the same after one generation's breeding, and therefore after any number. All

frequencies would thus diminish in the same proportion. Let λ stand for the fraction by which each frequency is multiplied, so that $u_1 = \lambda u_0$, $v_1 = \lambda v_0$, ..., $z_1 = \lambda z_0$. Then, substituting these as solutions of our six equations we find a set of six equations, involving λ, which are homogeneous and linear in the quantities $u_0/4$, $v_0/8$, etc., written as the column headings. For example the second line of the table gives

$$(2-4\lambda)\,\frac{u_0}{4} + 4 \cdot \frac{v_0}{8} + 0 \cdot \frac{w_0}{1} + 4 \cdot \frac{x_0}{16} + 0 \cdot \frac{y_0}{2} + 0 \cdot \frac{z_0}{4} = 0,$$

and the remaining five supply five more such equations.

The matrix of the coefficients of these equations is called the λ-matrix, and may be written down directly from the generation matrix; thus the λ-matrix for sib-matings is

TABLE 6

λ-matrix for Sib-matings

2−4λ	4	.	4	.	.
1	2−8λ	1	3	1	1
.	1	−λ	.	.	.
.	.	.	6−16λ	1	2
.	.	.	2	−2λ	.
.	.	.	.	.	1−4λ

The divisors of the columns now appear in the coefficients of $(-\lambda)$ in the diagonal.

Now the condition that six homogeneous linear equations in six unknowns shall be simultaneously satisfied is that the determinant of the matrix shall be zero. This determinant is of the sixth degree in λ, so that six roots are to be expected. Owing to the absence of entries in the lower blocks, the determinant of the

whole matrix is, in this case, the product of the deter-
minants of the blocks along the diagonal. There will
then be three equations, of the first, second and third
degrees respectively, in place of one of the sixth degree.
The roots of these equations are known as the latent
roots of the generation-matrix. They are the values for
which the λ-matrix becomes " singular," as matrices are
called when their determinants vanish.

From the bottom right-hand block we have the
linear equation

$$1-4\lambda = 0$$

with the root

$$\lambda = \tfrac{1}{4}.$$

From the middle block the equation is the quadratic

$$32\lambda^2 - 12\lambda - 2 = 0,$$

leading to the roots,

$$\lambda = \tfrac{1}{2}, \qquad \lambda = -\tfrac{1}{8}.$$

From the upper left-hand block, the determinant is
the cubic expression

$$-32\lambda^3 + 24\lambda^2 + 4\lambda - 2,$$

whence we have the equation

$$(4\lambda - 1)(4\lambda^2 - 2\lambda - 1) = 0,$$

of which the roots are

$$\lambda = \tfrac{1}{4}, \qquad \lambda = \tfrac{1}{4}(1 \pm \sqrt{5}).$$

The roots of algebraic equations may be positive or
negative, rational or irrational, real or imaginary. Of
the six roots obtained all are in this case real, two are
irrational involving the surd $\sqrt{5}$, and two are negative.
The largest root numerically, the one having the largest

absolute value, is called the dominant root.   In this case
it is positive and irrational, being

$$\tfrac{1}{4}(1+\sqrt{5}) = \cdot 80901,69944,$$

a value of rather frequent occurrence, which, for
convenience, we shall denote by the symbol, ϵ.

It will be noticed that two of the roots obtained
are equal to $\tfrac{1}{4}$; the value $\tfrac{1}{4}$ is then said to be a multiple
root, and for this value the λ-matrix is said to be
multiply singular.

10. Principal Components of Frequency

We have above derived the latent roots of the genera-
tion matrix from the condition necessary for the
frequencies of all mating types to change in proportion.
What these frequencies would have to be we have not
determined ; we do not know whether there are possible
non-negative frequencies at all.   A more fundamental
property of the latent roots comes into view when we
ask :—Is there any linear function of the frequencies
u . . . . z, such as

$$au+bv+cw+dx+ey+fz$$

which after one generation's breeding will in all cases
have fallen to the fraction λ of its former value ?   It
should be noticed that this property is not to be
conditional upon the mating types occurring in any
special frequency ratio.

As we have seen above z itself is one such linear
function, or principal component, of the frequencies,
corresponding with the multiple root $\lambda = \tfrac{1}{4}$, for, what-
ever may be the frequencies of the other mating types,

$$z_1 = \tfrac{1}{4}z_0.$$

To find such a linear function corresponding with
any latent root, we may substitute the value of the

root in the λ-matrix and read the *columns* of the matrix as homogeneous linear functions of the coefficients $a \ldots f$. In the case of the roots derived from the central block, we may ignore the equations derived from the first three columns, for these involve only a, b and c with coefficients forming a non-singular matrix, so that a, b and c must all be zero.

For $\lambda = \frac{1}{2}$, then we have the equations

$$\left.\begin{array}{r} -2d+2e = 0 \\ d - e = 0 \end{array}\right\} \text{Equivalent to one equation only}$$
$$2d - f = 0.$$

The equations derived from the columns passing through the block, the matrix of which has been made singular, are necessarily equivalent to one only, and with two homogeneous equations only among three unknowns their ratio is determinate, *i.e.*

$$\frac{d}{1} = \frac{e}{1} = \frac{f}{2},$$

so that the linear function,

$$L = x+y+2z,$$

or any multiple of it, possesses the property that, whatever may be the initial frequencies,

$$L_1 = \tfrac{1}{2}L_0,$$

and, therefore,

$$L_n = (\tfrac{1}{2})^n L_0.$$

Consequently, L, without regard to the arbitrary constant multiplier, is the principal component of the frequencies corresponding with the root $\lambda = \frac{1}{2}$.

Similarly, if we put $\lambda = -\frac{1}{8}$, we have the equations

$$\left.\begin{array}{r} 8d+2e = 0 \\ d+\tfrac{1}{4}e = 0 \end{array}\right\}$$
$$2d+\tfrac{8}{4}f = 0,$$

whence

$$\frac{d}{3} = \frac{e}{-12} = \frac{f}{-4},$$

so that, if

$$M = 3x - 12y - 4z,$$

any multiple of this linear function, possesses the property that

$$M_1 = -\tfrac{1}{8}.M_0$$

or

$$M_n = (-\tfrac{1}{8})^n M_0.$$

In order to gain a clear understanding of the procedure of solving equations derived from the columns of the λ-matrix, it is well to make up the functions L_1 and M_1 by multiplying the rows of the generation matrix by the corresponding coefficients and adding. The equations are designed to ensure that the original frequencies will then appear with the same coefficients multiplied by λ.

One immediate effect of recognising the principal components corresponding with the latent roots derived from the central block of the λ-matrix is to obtain explicit expressions for x and y in any future generation. For we can express x and y in terms of the principal components L, M and z.

Thus

$$x = \tfrac{1}{15}(12L + M - 20z)$$
$$y = \tfrac{1}{15}(3L - M - 10z) ;$$

so that

$$x_n = \tfrac{4}{5}L_n + \tfrac{1}{15}M_n - \tfrac{4}{3}z_n$$
$$= \tfrac{4}{5}(\tfrac{1}{2})^n L_0 + \tfrac{1}{15}(-\tfrac{1}{8})^n M_0 - \tfrac{4}{3}(\tfrac{1}{4})^n z_0,$$

while

$$y_n = \tfrac{1}{5}L_n - \tfrac{1}{15}M_n - \tfrac{2}{3}z_n$$
$$= \tfrac{1}{5}(\tfrac{1}{2})^n L_0 - \tfrac{1}{15}(-\tfrac{1}{8})^n M_0 - \tfrac{2}{3}(\tfrac{1}{4})^n z_0$$

and finally

$$z_n = (\tfrac{1}{4})^n z_0.$$

These equations are exact for all values of n. So far as they go, they enable us to construct the n-th power of the generation matrix, or the matrix appropriate to the lapse of n generations. It is more important that they pick out the terms of different orders of magnitude, for when n is large $(\frac{1}{8})^n$ or $(\frac{1}{4})^n$ will both be very small compared with $(\frac{1}{2})^n$. In fact x_n and y_n will be given to an increasingly close approximation by the simpler formulæ

$$x_n = \tfrac{4}{5}(\tfrac{1}{2})^n(x_0+y_0+2z_0)$$
$$y_n = \tfrac{1}{5}(\tfrac{1}{2})^n(x_0+y_0+2z_0).$$

The coefficients of the expression for L thus show how much trouble initial matings of types x, y and z will ultimately be in the elimination of matings with three genes, namely that matings of types x and y are equally troublesome, while a mating of type z is twice as bad as these, in the sense that it may be expected to contribute twice as much as they to the total frequencies of x_n and y_n as these become rare. It may also be noted that x tends to become four times as frequent as y, while z becomes very rare in comparison. Of the matings involving three or more genes still occurring, we may expect 80 per cent. to be of type x $(ab \times ac)$ and 20 per cent. of type y $(aa \times bc)$.

We may now pass to the roots belonging to the 3×3 block on the diagonal of the λ-matrix, and to consider first the value $\lambda = \frac{1}{4}$, which is the second appearance of this double root. The equations for the coefficients of the corresponding principal component are:—

$$\left.\begin{array}{l} a+b \ \ = 0 \\ 4a+c \ \ = 0 \\ b-\tfrac{1}{4}c = 0 \end{array}\right\}\text{Which are equivalent to two equations only}$$

and, from the other columns

$$\begin{array}{l} 4a+3b+2d+2e = 0 \\ b+ \ d-\tfrac{1}{2}e \ \ \ \ = 0 \\ b+2d+0\cdot f \ \ \ \ = 0. \end{array}$$

The value of f is indeterminate, so the coefficient of z in the principal component is arbitrary. This is a consequence of z being itself one of the principal components for this same root. For the rest we have

$$\frac{a}{2} = \frac{b}{-2} = \frac{c}{-8} = \frac{d}{1} = \frac{e}{-2},$$

so that in addition to z we have

$$P = 2u - 2v - 8w + x - 2y$$

as a principal component of the frequencies which in each generation has its value divided by 4. By combining this with an arbitrary multiple of z we have an infinitude of components satisfying the same condition. This set may, however, be compounded of any two of their number, if these are independent, that is, one is not a multiple of the other. In fact the λ-matrix, being doubly singular, may be premultiplied to zero not by a single row vector but by a two-row matrix

2	-2	-8	1	-2	0
0	0	0	0	0	1

which takes the place of the simple row-vector, as the principal component of frequencies corresponding with a multiple root.

For the remaining two roots, $\frac{1}{4}(1 \pm \sqrt{5})$, or ϵ and $\frac{1}{2} - \epsilon$, the algebraic manipulations are more complex in detail, although following the principles made clear in the simpler cases. The operations can, of course, be carried out arithmetically, putting in the values of the roots to any chosen accuracy. It is more satisfying to carry them out algebraically, and, since ϵ is a root of the quadratic equation $4\epsilon^2 - 2\epsilon - 1 = 0$, any polynomial in ϵ can be reduced to a linear form by substituting for ϵ^2, wherever it occurs, $\frac{1}{2}\epsilon + \frac{1}{4}$. Division by any

linear function of ϵ may be reduced to multiplication by observing that

$$(2\epsilon+r)(2\epsilon-r-1) = 1-r-r^2.$$

For the principal component corresponding with the root ϵ, we have the equations

$$
\left.
\begin{array}{ll}
(2-4\epsilon)a+b & = 0 \\
a+(2-8\epsilon)b+c & = 0 \\
b-\epsilon c & = 0
\end{array}
\right\}\text{Two independent equations}
$$
$$
\begin{array}{ll}
4a+3b+(6-16\epsilon)d+2e & = 0 \\
b+d-2\epsilon e & = 0 \\
b+2d+(1-4\epsilon)f & = 0.
\end{array}
$$

Hence, from the first equation

$$\frac{a}{1} = \frac{b}{4\epsilon-2}$$

but

$$(4\epsilon-2)\epsilon = 1\ ;$$

hence, from the third equation,

$$\frac{b}{4\epsilon-2} = \frac{\epsilon c}{4\epsilon-2} = \frac{c}{(4\epsilon-2)^2} = \frac{c}{8-8\epsilon}.$$

Eliminating e from the fourth and fifth equations we have

$$4\epsilon a+(3\epsilon+1)b+\{\epsilon(6-16\epsilon)+1\}d = 0,$$

whence

$$(3+2\epsilon)d = a\{4\epsilon+(3\epsilon+1)(4\epsilon-2)\} = a(8\epsilon+1),$$

or

$$d = \frac{a}{11}(8\epsilon+1)(4-2\epsilon) = 2\epsilon a.$$

Then, from the fifth equation,

$$e = (2\epsilon-1)(6\epsilon-2)a = (5-4\epsilon)a$$

and, from the seventh,

$$f = (4\epsilon-1)(8\epsilon-2)a = 2a.$$

TABLE

Matrix for n

	t_0	u_0	v_0	w_0
t_n	I	I $-\frac{1}{20}(5+16\epsilon)\epsilon^n$ 0 $-\frac{1}{20}(13-16\epsilon)(\epsilon')^n$ $-\frac{1}{10}\cdot 2^{-2n}$	I $-\frac{1}{10}(3+10\epsilon)\epsilon^n$ 0 $-\frac{1}{10}(8-10\epsilon)(\epsilon')^n$ $\frac{1}{10}\cdot 2^{-2n}$	I $-\frac{1}{20}(8+24\epsilon)\epsilon^n$ 0 $-\frac{1}{20}(20-24\epsilon)(\epsilon')^n$ $\frac{2}{5}\cdot 2^{-2n}$
u_n		$\frac{1}{5}(1+2\epsilon)\epsilon^n$ 0 $\frac{1}{5}(2-2\epsilon)(\epsilon')^n$ $\frac{2}{5}\cdot 2^{-2n}$	$\frac{4}{5}(\epsilon)\epsilon^n$ 0 $\frac{1}{5}(2-4\epsilon)(\epsilon')^n$ $-\frac{2}{5}\cdot 2^{-2n}$	$\frac{4}{5}\cdot\epsilon^n$ 0 $\frac{4}{5}(\epsilon')^n$ $-\frac{8}{5}\cdot 2^{-2n}$
v_n		$\frac{2}{5}(\epsilon)\epsilon^n$ 0 $\frac{1}{5}(1-2\epsilon)(\epsilon')^n$ $-\frac{1}{5}\cdot 2^{-2n}$	$\frac{2}{5}\cdot\epsilon^n$ 0 $\frac{2}{5}(\epsilon')^n$ $\frac{1}{5}\cdot 2^{-2n}$	$\frac{1}{5}(8\epsilon-4)\epsilon^n$ 0 $-\frac{8}{5}(\epsilon)(\epsilon')^n$ $\frac{4}{5}\cdot 2^{-2n}$
w_n		$\frac{1}{20}\cdot\epsilon^n$ 0 $\frac{1}{20}\cdot(\epsilon')^n$ $-\frac{1}{10}\cdot 2^{-2n}$	$\frac{1}{10}(2\epsilon-1)\epsilon^n$ 0 $-\frac{1}{5}(\epsilon)(\epsilon')^n$ $\frac{1}{10}\cdot 2^{-2n}$	$\frac{1}{5}(2-2\epsilon)\epsilon^n$ 0 $\frac{1}{5}(1+2\epsilon)(\epsilon')^n$ $\frac{2}{5}\cdot 2^{-2n}$
x_n				
y_n				
z_n				

NOTE.—ϵ' is

7

Generations Sib-mating

x_0	y_0	z_0
I	I	I
$-\frac{1}{10}(4+13\epsilon)\epsilon^n$	$-\frac{1}{20}(9+28\epsilon)\epsilon^n$	$-\frac{1}{10}(5+16\epsilon)\epsilon^n$
$\frac{1}{2}.2^{-n}$	$\frac{1}{2}.2^{-n}$	2^{-n}
$-\frac{1}{20}(21-26\epsilon)(\epsilon')^n$	$-\frac{1}{20}(23-28\epsilon)(\epsilon')^n$	$-\frac{1}{10}(13-16\epsilon)(\epsilon')^n$
$-\frac{1}{20}.2^{-2n}$	$\frac{1}{10}.2^{-2n}$	$-\frac{1}{5}.2^{-2n}$
$\frac{1}{5}(1+4\epsilon)\epsilon^n$	$\frac{1}{5}(3+2\epsilon)\epsilon^n$	$\frac{1}{5}(2+4\epsilon)\epsilon^n$
-2^{-n}	-2^{-n}	-2.2^{-n}
$\frac{1}{5}(3-4\epsilon)(\epsilon')^n$	$\frac{1}{5}(4-2\epsilon)(\epsilon')^n$	$\frac{1}{5}(4-4\epsilon)(\epsilon')^n$
$\frac{1}{5}.2^{-2n}$	$-\frac{2}{5}.2^{-2n}$	$\frac{4}{5}.2^{-2n}$
$\frac{1}{5}(1+2\epsilon)\epsilon^n$	$\frac{1}{5}(6\epsilon-2)\epsilon^n$	$\frac{4}{5}(\epsilon)\epsilon^n$
$-\frac{2}{5}.2^{-n}$	$-\frac{2}{5}.2^{-n}$	$-\frac{1}{5}.2^{-n}$
$\frac{1}{5}(2-2\epsilon)(\epsilon')^n$	$\frac{1}{5}(1-6\epsilon)(\epsilon')^n$	$\frac{1}{5}(2-4\epsilon)(\epsilon')^n$
$-\frac{1}{10}.2^{-2n}$	$\frac{1}{5}.2^{-2n}$	$\frac{4}{15}.2^{-2n}$
$-\frac{1}{10}(-2)^{-3n}$	$\frac{2}{5}(-2)^{-3n}$	$\frac{2}{15}(-2)^{-3n}$
$\frac{1}{10}(\epsilon)\epsilon^n$	$\frac{1}{20}(5-4\epsilon)\epsilon^n$	$\frac{1}{10}.\epsilon^n$
$-\frac{1}{10}.2^{-n}$	$-\frac{1}{10}.2^{-n}$	$-\frac{1}{5}.2^{-n}$
$\frac{1}{20}(1-2\epsilon)(\epsilon')^n$	$\frac{1}{20}(3+4\epsilon)(\epsilon')^n$	$\frac{1}{10}(\epsilon')^n$
$-\frac{1}{20}.2^{-2n}$	$\frac{1}{10}.2^{-2n}$	$\frac{2}{15}.2^{-2n}$
$\frac{1}{10}(-2)^{-3n}$	$-\frac{2}{5}(-2)^{-3n}$	$-\frac{2}{15}(-2)^{-3n}$
$\frac{4}{5}.2^{-n}$	$\frac{4}{5}.2^{-n}$	$\frac{8}{5}.2^{-n}$
o	o	$-\frac{4}{3}.2^{-2n}$
$\frac{1}{5}(-2)^{-3n}$	$-\frac{4}{5}(-2)^{-3n}$	$-\frac{4}{15}(-2)^{-3n}$
$\frac{1}{5}.2^{-n}$	$\frac{1}{5}.2^{-n}$	$\frac{2}{5}.2^{-n}$
o	o	$-\frac{2}{3}.2^{-2n}$
$-\frac{1}{5}(-2)^{-3n}$	$\frac{4}{5}(-2)^{-3n}$	$\frac{4}{15}(-2)^{-3n}$
		2^{-2n}

written for $\frac{1}{2}-\epsilon$.

The required principal component corresponding with the root ϵ is therefore

$$Q = u + (4\epsilon - 2)v + (8 - 8\epsilon)w + 2\epsilon x + (5 - 4\epsilon)y + 2z,$$

and the corresponding component for the negative root $\frac{1}{2} - \epsilon$ is found by substituting $\frac{1}{2} - \epsilon$ for ϵ, and is

$$R = u - 4\epsilon v + (4 + 8\epsilon)w + (1 - 2\epsilon)x + (3 + 4\epsilon)y + 2z.$$

Since Q is the component corresponding with the dominant root, the coefficients are all positive, and represent the extent to which different types of initial mating will contribute to the heterozygosity which remains when all other components have been reduced, relative to the dominant component, to negligible quantities.

To the six latent roots, therefore, there correspond six linear components of frequency each of which diminishes from generation to generation in a fixed ratio. The frequency of any mating type after n generations may be expressed in terms of these, and thus calculated from the initial values.

Essentially this requires no more than the expressions giving the frequency of each mating type in terms of the principal components, *i.e.*

$$z = \quad . \quad\quad . \quad\quad . \quad\quad . \quad\quad . \quad\quad z$$

$$y = \quad . \quad\quad . \quad\quad . \quad\quad . \quad \frac{1}{5}L - \frac{1}{15}M - \frac{2}{3}z$$

$$x = \quad . \quad\quad . \quad\quad . \quad\quad . \quad \frac{4}{5}L + \frac{1}{15}M - \frac{4}{3}z$$

$$w = -\frac{1}{20}P + \frac{1}{20}Q + \frac{1}{20}R - \frac{1}{10}L + \frac{1}{30}M + \frac{2}{15}z$$

$$v = -\frac{1}{10}P + \frac{2\epsilon}{5}Q + \frac{1-2\epsilon}{5}R - \frac{2}{5}L - \frac{1}{30}M + \frac{4}{15}z$$

$$u = \quad \frac{1}{5}P + \frac{1+2\epsilon}{5}Q + \frac{2-2\epsilon}{5}R - L \quad\quad + \frac{4}{5}z.$$

These equations are valid for all generations. To find the frequencies after n generations, we have merely to express the principal components after n generations in terms of their initial value, by such relations as

$$P_n = (\tfrac{1}{4})^n P_0$$
$$Q_n = \epsilon^n Q_0, \text{ etc.}$$

The condition of the system is thus expressed in terms of quantities known from the original condition. If, further, the initial values of the principal components are replaced by their expressions in terms of the initial frequencies of the different mating types, we shall develop the n-th power of the generation matrix, or the transformation matrix for n generations.

11. The Frequency Distributions Corresponding with the Latent Roots

In section 3 the notion of latent roots was introduced with the supposition that, for some special distribution of frequency, the frequencies of all mating types might decrease in proportion. It is now seen that such distributions can be found by equating to zero all five linear components other than the one corresponding with the root concerned. For example, if z, L, M, P and R were all zero, we should have five equations determining the relative frequencies of the mating types u . . . . z, and, as in subsequent generations all these components continue to be zero, and the surviving component diminishes from generation to generation in the ratio ϵ, it is clear that this will be the right frequency ratio to make all frequencies decline proportionately in this ratio. Evidently x, y and z will be zero, and the remaining three frequencies may be found from the equations

$$P = 2u - 2v - 8w = 0$$
$$R = u - 4\epsilon v + (4 + 8\epsilon)w = 0.$$

From these equations it appears that

$$\frac{u}{8+48\epsilon} = \frac{v}{16+16\epsilon} = \frac{w}{8\epsilon-2} = \frac{u+v+w}{22+72\epsilon};$$

but

$$(22+72\epsilon)(36\epsilon-29) = 10,$$
$$(8+48\epsilon)(36\epsilon-29) = 200 - 240\epsilon,$$
$$(16+16\epsilon)(36\epsilon-29) = 400\epsilon - 320,$$

and

$$(8\epsilon- 2)(36\epsilon-29) = 130 - 160\epsilon,$$

so that the frequency ratio, written in terms of $u+v+w$, becomes

$$\frac{u}{20-24\epsilon} = \frac{v}{40\epsilon-32} = \frac{w}{13-16\epsilon} = \frac{u+v+w}{1}.$$

The frequency distribution corresponding with the dominant root is best expressed in terms of the total frequency, for it represents the actual frequency distribution of those matings which are not yet between like homozygotes, to which the inbred line will tend as all components other than the dominant component become negligible. If we insert the value of ϵ we then find

Mating Type.	Symbol.	Frequency.	Per cent.
$aa \times ab$	u	$20 - 24\epsilon$	58·359
$ab \times ab$	v	$40\epsilon-32$	36·068
$aa \times bb$	w	$13 - 16\epsilon$	5·573
		1	100·000

and these will be the proportionate frequencies of these mating types at loci where complete homogeneity has not yet been established.

Moreover if these frequencies are substituted in the linear function Q, we find

$$u = 20- 24\epsilon$$
$$(4\epsilon-2)v = (4\epsilon-2)(40\epsilon-32) = 104-128\epsilon$$
$$(8-8\epsilon)w = (8-8\epsilon)(13-16\epsilon) = 136-168\epsilon$$

$$\overline{ 260-320\epsilon}$$

but

$$(260 - 320\epsilon)(5 + 16\epsilon) = 20$$

hence, if we multiply Q by the factor $(5 + 16\epsilon)/20$ and obtain

$$(5 + 16\epsilon)u + (6 + 20\epsilon)v + (8 + 24\epsilon)w + (8 + 26\epsilon)x + (9 + 28\epsilon)y +$$
$$(10 + 32\epsilon)z \div 20$$

we have an expression which, multiplied by ϵ^n, gives the absolute frequency of non-homogeneous matings n generations later. The coefficients of the frequencies in this expression measure the complexity of each mating type relative to the average of the types to which the population tends. They may be represented by like capital letters corresponding with those used for the frequencies. Thus, we have the following values

Mating Type.		Complexity, in Standard Measure.	
$aa \times ab$	U	$5 + 16\epsilon$	·8972136
$ab \times ab$	V	$6 + 20\epsilon$	1·1090170
$aa \times bb$	W	$8 + 24\epsilon$	1·3708204
$ab \times ac$	X	$8 + 26\epsilon$	1·4517221
$aa \times bc$	Y	$9 + 28\epsilon$	1·5826238
$ab \times cd$	Z	$10 + 32\epsilon$	1·7944272
		$\div 20$	

The principle of standardisation of the two series, the frequencies and the complexities corresponding with the dominant latent root, is that the sum of the frequencies should be equal to unity, and the sum of the products of the frequencies with the complexities should also be equal to unity. Thus

$$u + v + w = 1$$
$$uU + vV + wW = 1.$$

The complexities of each mating type will then measure the value for each, which when multiplied by ϵ^n gives its contribution to the probability of a later mating being still heterogeneous.

In other problems we may require the frequencies corresponding with the dominant root without having worked out the other roots and the corresponding principal components. This may be done by treating the *rows* of the λ-matrix as equations in u, v and w. Using the rows, the divisors of the columns can no longer be ignored. So, for the root ϵ, we have the equations

$$(2-4\epsilon)\frac{u}{4} + 4 \cdot \frac{v}{8} = 0$$

$$\frac{v}{8} - \epsilon \cdot w = 0,$$

whence

$$\frac{u}{2} = \frac{v}{4\epsilon-2} = \frac{8w}{(4\epsilon-2)^2} = \frac{w}{1-\epsilon} = \frac{u+v+w}{1+3\epsilon},$$

but

$$(1+3\epsilon)(10-12\epsilon) = 1,$$

so

$$u = 20 - 24\epsilon$$
$$v = 40\epsilon - 32$$
$$w = 13 - 16\epsilon$$

is the solution adding to unity, agreeing with that obtained by the method first used.

The frequency series corresponding with any simple root may thus be recognised as a column vector by which the λ-matrix is post-multiplied to zero, just as the complexity series is a row vector by which it is premultiplied to zero. In the case of multiple roots these vectors are replaced by matrices of two or more columns or rows respectively.

12. The Amount of Inbreeding to which Material has been Subjected

The rate of progress towards homozygosity depends essentially on the dominant latent root, for, if the different mating types are scored according to their

respective complexities, the total score will diminish in the ratio ϵ in each generation, irrespective of the frequency of different mating types. To compare the rates appropriate to different mating systems we may compare the values of $(-\log_e \lambda)$, or of $\log_e \frac{1}{\lambda}$, for the dominant roots. For example, when self-fertilisation can be practised the proportion of heterozygotes will, in the absence of selection, be halved in each generation. Since after n generations the proportion is diminished in the ratio λ^n, a sib-mated line will have received an equal amount of inbreeding to a selfed line if the number of generations of sib-mating exceeds the number of generations of selfing in the ratio $\log 2/(-\log \epsilon)$. For any method of inbreeding the number of generations required to make a fixed amount of progress may be measured by $1/(-\log \lambda)$. Thus

							Selfing.	Sib-mating.
λ	.	.	.	.	.	.	·5	·809017
$-\log \lambda$	.	.	.	.	.	.	·69315	·21194
Equivalent number of generations				.			1·443	4·718

The ratio of the logarithms, which gives the relative speed of the two processes is $3 \cdot 2705$. An absolute measure of the amount of progress made in inbreeding is found by multiplying the number of generations by $\log 1/\lambda$. This is a practical measure of what has been done. When it reaches 2 units a great deal of progress has been made ; when it reaches 5 units homozygosis is nearly complete. This takes 7 to 8 generations of selfing, or 23 to 24 generations of sib-mating. This measure of the amount of progress made does not, of course, measure the actual degree of homozygosis attained, for the extent to which the germ plasm was homogeneous in the original material is always unknown. It is this which renders attempts to set up coefficients of inbreeding unsatisfactory.

We do know, however, that no locus can be in worse condition than the mating z, with complexity relative to standard of 1·7944. The natural logarithm of this is ·5847, so that, using the standard condition as basis, no locus can be more than ·5847 behind reckoning; this is equivalent to 2·759 generations. As matings involving four allelomorphs may be thought to be uncommon, and as it is not improbable that a considerable fraction of the germ-plasm may be homogeneous from the start, it will usually be convenient to make comparisons on the assumption that all loci were originally in a heterogeneous condition at the standard level of complexity. A fairly common initial condition must be type w, $aa \times bb$, of which the complexity is 1·3708 times the standard, or ·3154 logarithmic units behind reckoning. Unless, however, loci in this condition were nearly three times as frequent as loci initially homogeneous, the frequency of heterogeneity in later generations will not be underestimated.

These conclusions are not immediately applicable to certain special cases. Of these we may note loci on the pairing and differential segments of the sex chromosome, loci homologous with those of genes kept segregating, and loci closely linked in the same chromosome with genes kept segregating. These cases require separate investigation which will be carried out in the following chapters.

The amount of inbreeding to which a sib-mating line has so far been subjected is, therefore, conveniently measured by multiplying the number of generations by ·212, or more accurately ·21194. On looking back over the genealogy, however, we may find it to consist of two or more sections of regular sib-matings broken by an irregularity or departure from the regular process. For example, a female may have been mated to her father, as often seems advantageous when there is no

male in the first litter, or a male may have been mated to a half-sister born to his father by a sister of his parents. In such cases, before we can evaluate the degree of inbreeding indicated by the pedigree, we shall need to find how much, if anything, the irregular portion has added to the number of generations in the regular portion. Such an evaluation is useful in another way, for, at the expense of sufficient time it would always be possible to adhere strictly to a given mating programme. The irregularities have in fact no doubt taken place under the impression that the additional time needed for the regular series of sib-matings would lead at a given future date to the available material being less inbred than if, for the sake of expedition, an irregularity were introduced. Both for the interpretation of records, therefore, and for practical use in avoiding delay in the progress towards homogeneity, we require to be able to evaluate an irregularity in the lineage, of any particular kind.

13. The Evaluation of an Irregularity

Tracing an inbred pedigree back we may pass through a number of sib-matings but, unless the series is wholly regular, shall reach a stage at which a mating of some other kind was made, after which the line has been maintained wholly by sib-matings. This is called the last irregular mating, or the last mating of the irregularity. The complete ancestry of the two animals used in this mating is then traced back until all lines of ascent are united in a single mating. This, if a sib-mating, is the last regular mating before the irregularity. The irregularity itself consists in the transition from the last regular mating to the last irregular mating. We have the means of evaluating the irregularity, of whatever kind it may be, when we

know with what frequency each possible type of last regular mating leads to each possible type of last irregular mating. We thus construct a matrix for the irregularity.

For example, if the irregularity consists in intercalating a single mating of parent and offspring in a sib-mating series, there is only one irregular mating. Following out the consequences of the three types $aa \times ab$, $ab \times ab$ and $aa \times bb$ of last regular mating we find

TABLE 8

Matrix of Intercalated Parent-offspring Mating

	$aa \times ab$	$ab \times ab$	$aa \times bb$
$aa \times aa$	$\frac{1}{4}$	·	·
$aa \times ab$	$\frac{1}{2}$	$\frac{1}{4}$	1
$ab \times ab$	$\frac{1}{4}$	$\frac{1}{2}$	·
$aa \times bb$	·	·	·

as the matrix for the irregularity. Now, if such an irregularity has been preceded by a number of sib-matings, we know the relative frequencies u, v and w of the possible mating types at the heads of the columns. By multiplying each column by u, v or w and adding we therefore have the expected frequencies of all possible mating types for the last irregular mating. If these are multiplied by the complexity of each and added, we have a measure of the complexity of the material from which the new series of sib-matings is derived, and so a measure of the effect of the irregularity. Since the mating $aa \times aa$ has zero complexity, the first row of the matrix may be omitted. From the other rows, inserting the expressions for u, v and w, we have

$$aa \times ab \qquad 10-12\epsilon+20\epsilon-16+13-16\epsilon = \quad 7 - 8\epsilon$$
$$ab \times ab \qquad 5- 6\epsilon+20\epsilon-16 \qquad\quad = 14\epsilon-11$$
$$aa \times bb \qquad\qquad\qquad\qquad\qquad\qquad = \quad 0.$$

Multiplying these rows by their complexities and adding, we find

$$\tfrac{1}{20}\{(7-8\epsilon)(5+16\epsilon)+(14\epsilon-11)(6+20\epsilon)\}$$
$$= \tfrac{1}{20}\{3+8\epsilon+4+4\epsilon\}$$
$$= \tfrac{1}{20}\{7+12\epsilon\} = \cdot 83541.$$

This is less than unity, so that the irregular mating has been responsible for some progress. It is not so small as the ratio for a regular mating, ·80902, so that it is not worth a full generation of sib-mating, and a sib-mating, if one could have been made equally soon, should have been preferred. The absolute progress made by the irregularity is—log (·83541) = ·17985, equivalent to the fraction ·84852 of a normal generation. It has thus been advantageous if the time saved by making it was as great as one-sixth of the average generation time.

Because a mating of parent and offspring interpolated in a series of sib-matings does not advance the inbreeding process so much as a sib-mating would have done, it should not be argued that continuous series of parent-and-offspring matings would be slower than a series of sib-matings. This is not the case. Parent-offspring inbreeding is equally efficient with sib-inbreeding, as will be shown in the following chapter, and for particular purposes is certainly to be preferred. Moreover, a sib-mating interpolated in a series of parent-offspring matings retards progress to exactly the same extent (·15148 generation) as does a parent-offspring mating interpolated in a series of sib-matings. A mixture of the two types is less satisfactory than either taken alone.

Consider now a case where the irregularity con-sists of three matings. A brother is mated to two

whole sisters, and offspring from the two unions are interbred.

If the last regular mating were of type u $(aa \times ab)$, the possible pairs of derived matings are

$$\text{I } tt \quad 2 \text{ } tu \quad 2 \text{ } uu \quad 2 \text{ } uv \quad \text{I } vv, \text{ out of 8 ;}$$

from a mating of type v $(ab \times ab)$ we shall obtain

$$\text{I } tt \text{ } 4 \text{ } tu \text{ } 2 \text{ } uu \text{ } 2 \text{ } tw \text{ } 4 \text{ } uw \text{ } \text{I } ww \text{ } 2 \text{ } uu' \text{ } 8 \text{ } uv \text{ } 4 \text{ } vv, \text{ out of 32 ;}$$

while from type w $(aa \times bb)$ we can only obtain

$$vv.$$

Here uu' stands for the pair of matings $aa \times ab$, $ab \times bb$.

Matings between the young, one from each mating, of these pairs of matings are

<div align="center">

TABLE 9

Types of Parallel Matings

			t	u	v	w	Divisor.
tt	.	.	I	.	.	.	I
tu	.	.	I	I	.	.	2
uu	.	.	I	2	I	.	4
uv	.	.	I	4	2	I	8
vv	.	.	I	4	2	I	8
tw	.	.	.	I	.	.	I
uw	.	.	.	I	I	.	2
ww	.	.	.	.	I	.	I
uu'	.	.	.	2	I	I	4

</div>

whence the matrix for the irregularity can be constructed

<div align="center">

TABLE 10

Matrix for Intercalated Half-sib-Mating

				$\dfrac{u}{64}$	$\dfrac{v}{16}$	$\dfrac{w}{8}$
t	.	.	.	23	3	I
u	.	.	.	28	8	4
v	.	.	.	10	4	2
w	.	.	.	3	I	I

</div>

Ignoring the first line, and scoring the columns with the known values of u, v, w we find

$$\frac{U}{64}\left\{560-672\epsilon-1024+1280\epsilon+416-512\epsilon\right\} = \frac{U}{4}(6\epsilon-3)$$

$$\frac{V}{64}\left\{200-240\epsilon-512+640\epsilon+208-256\epsilon\right\} = \frac{V}{8}(18\epsilon-13)$$

$$\frac{W}{64}\left\{60-72\epsilon-128+160\epsilon+104-128\epsilon\right\} = \frac{W}{16}(9-10\epsilon),$$

whence, substituting for U, V and W there remains

$$\tfrac{1}{80}\{(6\epsilon-3)(5+16\epsilon)+(18\epsilon-13)(3+10\epsilon)+(9-10\epsilon)(2+6\epsilon)\}$$
$$= \tfrac{1}{80}\cdot(9+30\epsilon+6+14\epsilon+3+4\epsilon)$$
$$= \tfrac{1}{40}\cdot(9+24\epsilon)$$
$$= \cdot71041$$

which is the ratio of reduction effected in two generations. The logarithm is $\cdot34193$, equivalent to $1\cdot613$ normal generations. A mating of this type would thus require a considerable advantage in time, equivalent to $\cdot387$ of a generation, before it could profitably be used.

If offspring from two parallel sib-matings having no member in common were used, the ratio of reduction comes to $\tfrac{1}{80}(18+55\epsilon) = \cdot78120$, equivalent to only $1\cdot1651$ normal generations. Such a mating can seldom be the best course available. Nevertheless, as it may have been used, it is desirable to be able to assess its contri-bution to the whole course of inbreeding.

The types of irregularity possible are somewhat numerous, but the greater part of the labour of working out any new type lies in following out its genetic consequences so as to construct the corresponding table of frequencies. Multiplying the columns by $u, v, \ldots$ and the rows by $U, V, \ldots$ is a comparatively simple process.

14. Lengths of Tracts of Heterogeneous Origin

If a dominant gene is carried in an inbred line, with three recessives at the same locus, it will be

associated with a tract of chromosome on either side, introduced with the dominant gene. The length of each of these tracts may diminish, but cannot increase. The frequency distribution of the length of each tract after n generations may be found as follows :—

To every point on the same chromosome as the dominant gene, there will be a certain probability s that one or more breaks shall occur between this point and the locus of the dominant. The probability that no break occurs in n generations is, therefore,

$$(1-s)^n.$$

When n is large this is negligible, save when s is so small that sn is not large, and for such small values of s the expression tends to equivalence with

$$e^{-sn}.$$

When s is small, e.g. 0·1 or less, this quantity is indistinguishable from two others between which it lies, (a) the recombination frequency, and (b) the map distance. Consequently, if x is the map distance of any locus from the fixed dominant, the probability that the intervening tract is unbroken after n generations is

$$p = e^{-nx},$$

and the frequency of the map length x of the attached unbroken tract is $df = e^{-nx} n\, dx.$

The average length of the attached tract is

$$\int x\, df = \frac{1}{n} \int_0^a (nx) e^{-nx} n\, dx = \frac{1}{n} = \frac{100}{n} \text{ centimorgans.}$$

There will be two such tracts associated on its two sides with each dominant gene selected.

The question of the lengths of tracts of heterogeneous origin in this and other parts of the germ-plasm

requires a somewhat more elaborate examination. The argument developed above may be used to show that the unbroken segments of each of the four homologous chromosomes in any mating are of average length $100/n$, but the tracts of heterogeneous origin should not be identified with such unbroken segments, for one end may be determined by a break in one, and the other end by a break in any one of the three other chromosomes. Consequently, the average length of such a tract might be expected to be shorter than $100/n$. This, as will be seen, is true. The same breaks must, however, often affect different chromosomes, and many of them will be ineffective. Some will occur in individuals already homozygous in the region of breakage. Others which have formed junctions between strands of unlike origin may quickly disappear through chance segregation. In other cases matings will become homogeneous for the junction formed, so that the junction remains as the point of division of segments in both of which the line is homogeneous. If one segment is homogeneous but not the other, the junction will form the terminus of a heterogeneous tract ; if both segments are heterogeneous, it will lie within such a tract.

If the total map length of the species is L, the total map length still heterogeneous in the n-th sib-mating is the coefficient of z_0 in $1 - t_n$, multiplied by L, $i.e.$

$$\{\tfrac{1}{10}(5 + 16\epsilon)\epsilon^n - (\tfrac{1}{2})^n + \tfrac{1}{10}(13 - 16\epsilon)(\tfrac{1}{2} - \epsilon)^n + \tfrac{1}{5}(\tfrac{1}{4})^n\}L$$

of which the leading term is

$$LZ\epsilon^n$$

where Z is about $1 \cdot 794$.

The rate of production of junctions between segments of unlike origin in gametogenesis from the r-th generation will be the coefficient of z_0 in

$$2u_r + 4v_r + 4x_r + 2y_r + 4z_r = \tfrac{1}{5}(2 + 12\epsilon)Q_r + \tfrac{1}{5}(8 - 12\epsilon)R_r$$

multiplied by L, since in matings of type w both parents

are homozygous, and in types u and y one is homozygous, so that two only of the four new chromosomes formed are derived from heterozygotes. This gives

$$\tfrac{1}{6}(4+24\epsilon)\epsilon^r+\tfrac{1}{6}(16-24\epsilon)(\tfrac{1}{2}-\epsilon)^r.$$

Now, when a junction occurs, one but not the other of the members of the next $(r+1)$-th mating may be spoken of as heterozygous for the junction, which will be transmitted as though it were a point mutation. In respect of the junction therefore the $(r+1)$-th mating will be of type u, and the coefficient of u_0 in

$$u_{n-r-1}+v_{n-r-1}+w_{n-r-1}$$

will be the probability of junctions formed in gametogenesis from the r-th generation being not homozygous that is being bounded by one or two heterogeneous segments, in the n-th generation. The total number of such junctions accumulated in the n-th generation will be found by summing for all values of r, *i.e.* for all the stages at which junctions can originate. This is

$$L.\sum_{r=0}^{n-1}\left\{\left[\frac{1}{20}(5+16\epsilon)\epsilon^{n-r-1}+\frac{13-16\epsilon}{20}(\tfrac{1}{2}-\epsilon)^{n-r-1}+\frac{1}{10}(\tfrac{1}{4})^{n-r-1}\right]\right.\\ \left.\times[\tfrac{1}{6}(4+24\epsilon)\epsilon^r+\tfrac{1}{6}(16-24\epsilon)(\tfrac{1}{2}-\epsilon)^r].\right\}$$

For the simplification of this and similar expressions we may note that

$$\sum_{r=0}^{n-1}\alpha^r\beta^{n-r-1}=\frac{1}{\alpha-\beta}(\alpha^n-\beta^n)\text{ where }\alpha\neq\beta,$$
$$\text{or }n\alpha^{n-1}\qquad\text{where }\alpha=\beta.$$

The leading term is therefore

$$\tfrac{1}{100}Ln\ (4+24\epsilon)(5+16\epsilon)\epsilon^{n-1}$$
$$=\tfrac{1}{100}Ln\ (116+376\epsilon)\epsilon^{n-1}$$
$$=\tfrac{1}{100}Ln\ (144+464\epsilon)\epsilon^{n}$$
$$=\tfrac{1}{6}LZn(16+16\epsilon)\epsilon^{n}.$$

Hence, subject to correction by terms of lower order, and by allowance for internal junctions, and allowing two termini to each fragment, the average number of tracts per unit crossover length of the heterogeneous portion of the chromosome is approximately

$$\frac{n}{10}(8+8\epsilon) = 1\cdot44721\ n,$$

and a typical length of each tract is

$$\frac{1000}{(8\epsilon+8)n} = \frac{50(3-2\epsilon)}{n} = \frac{69\cdot0983}{n}\ \text{centimorgans.}$$

The remaining terms involving $L\epsilon^n$ are

$$\frac{1}{100}(4+24\epsilon)(13-16\epsilon)\frac{1}{2\epsilon-\frac{1}{2}} = \frac{Z\epsilon^n}{5}(356-440\epsilon)$$

$$\frac{1}{50}(4+24\epsilon)\frac{1}{\epsilon-\frac{1}{4}} = \frac{Z\epsilon^n}{5}(80-96\epsilon)$$

$$\frac{1}{100}(5+16\epsilon)(16-24\epsilon)\frac{1}{2\epsilon-\frac{1}{2}} = \frac{Z\epsilon^n}{5}(-8+8\epsilon)$$

or in all

$$\tfrac{1}{5}(428-528\epsilon)Z\epsilon^n$$

so that, to this degree of approximation, the average density of junctions is

$$\tfrac{1}{5}\{(8+8\epsilon)n+(428-528\epsilon)\}.$$

The frequency of junctions within heterogeneous tracts remains to be subtracted. The analogy of a junction with a mutation may again be used, only now we shall consider the frequency, not of matings heterogeneous for the junction, but of those which beside being heterogeneous for a junction involve chromosomes, not involving the junction, of two different kinds. These will be matings of type x, y or z. Half of the matings derived from type u, in the gametogenesis of which a junction occurs, will be of type x; from type

v one half are of type x and a quarter of type y. We shall need, therefore, in one factor the leading term in the coefficient of z_0 in

$$\tfrac{1}{2}(2u)+\tfrac{3}{4}(4v) = u+3v$$

or

$$.\tfrac{1}{5}(2+16\epsilon)\epsilon^r.$$

Type u ————————— ················· × ················· —————————
················· ·················

leads to

type u ————————— ·········· × ················· —————————
················· ·················

or to

type x ————————— ·········· × ·········· —————————
················· ·· ·················

Type v ————————— ················· × ················· —————————
················· ·················

leads to

type u ————————— ·········· × ————————— —————————
————————— —————————

or

type x ————————— ·········· × ————————— —————————
————————— ·················

or

type y ————————— ·········· ················· × ················· —————————
————————— ·················
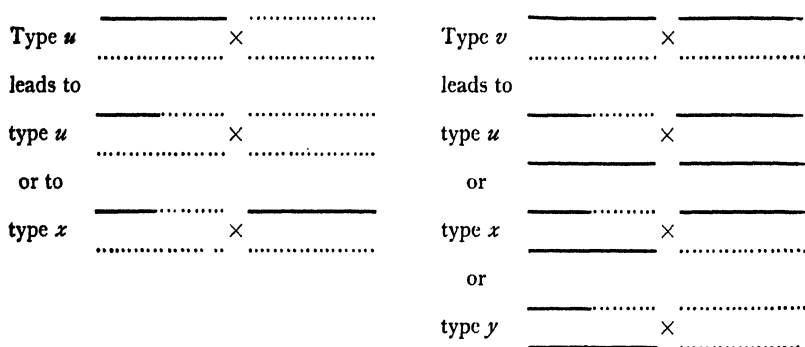
FIG. 2.—Modes of generating junctions internal to heterogenic regions.

Next, observe that the coefficient both of x_0 and y_0 in

$$x_{n-r-1}+y_{n-r-1}+z_{n-r-1}$$

has the value

$$(\tfrac{1}{2})^{n-r-1}$$

But

$$\sum_{r=0}^{n-1} \tfrac{1}{5}(2+16\epsilon)\epsilon^r(\tfrac{1}{2})^{n-r-1}$$
$$= \tfrac{1}{5}(2+16\epsilon)\,\frac{1}{\epsilon-\tfrac{1}{2}}\,[\epsilon^n-(\tfrac{1}{2})^n],$$

so that the coefficient of ϵ^n is

$$\tfrac{1}{5}(2+16\epsilon)4\epsilon = \tfrac{1}{5}(16+40\epsilon) = Z(96-112\epsilon) = 5\cdot39010\,Z,$$

and the average density of internal junctions is only $5\cdot39010$ per 100 crossover units of the length still heterogeneous, or one in $18\cdot55$ units.

Deducting these from the total density of junctions associated with the heterogeneous region, there remain

$$\tfrac{1}{6}\{(8+8\epsilon)n-(52-32\epsilon)\} = \tfrac{1}{6}(8+8\epsilon)\{n-\tfrac{1}{8}(43-42\epsilon)\}.$$

Numerically this is

$$2\cdot89443n-5\cdot22229 = 2\cdot89443(n-1\cdot80426).$$

A, B, C, D are loci maintained in constant segregation
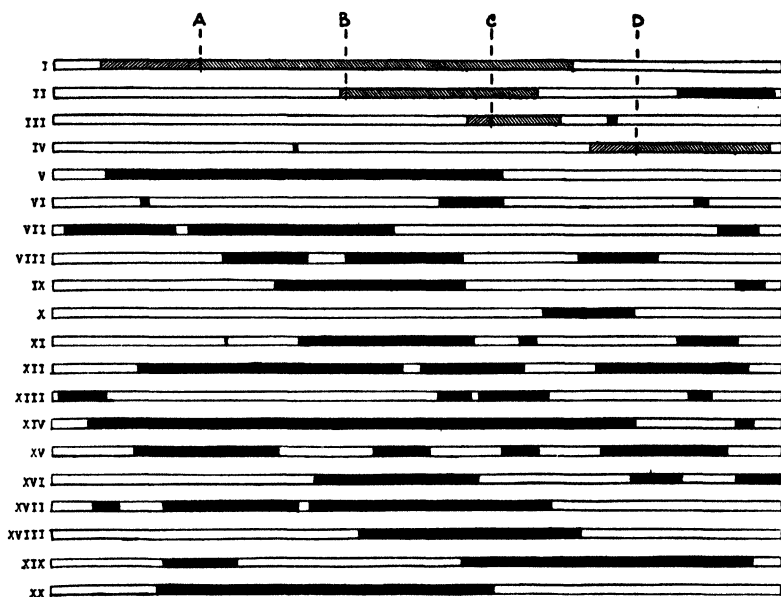
FIG. 3.—Model map of diploid organism with 20 chromosomes after 5 generations of sib-mating.

The number of separate heterogeneous tracts per unit length will be half of this, or

$$1\cdot44721(n-1\cdot80),$$

and the typical length may be taken to be approximately

$$\frac{69}{n-1\cdot8} \text{ centimorgans,}$$

when n considerably exceeds 2.

A distinction may be drawn between end tracts terminated by the end of a chromosome, and enclosed tracts. The former will be on the average only half the length of the latter. From this point of view the regions still possibly heterogeneous after n generations of sib-mating may be recognised as

	Number.	Average Length.
End tracts	$2\nu Z\epsilon^n$	$34 \cdot 5/(n - 1 \cdot 8)$
Enclosed tracts . . .	$\{(1 \cdot 45n - 2 \cdot 61)L - \gamma\}Z\epsilon^n$	$69/(n - 1 \cdot 8)$

where ν stands for the haploid number of chromosomes.

A, B, C, D are loci maintained in constant segregation
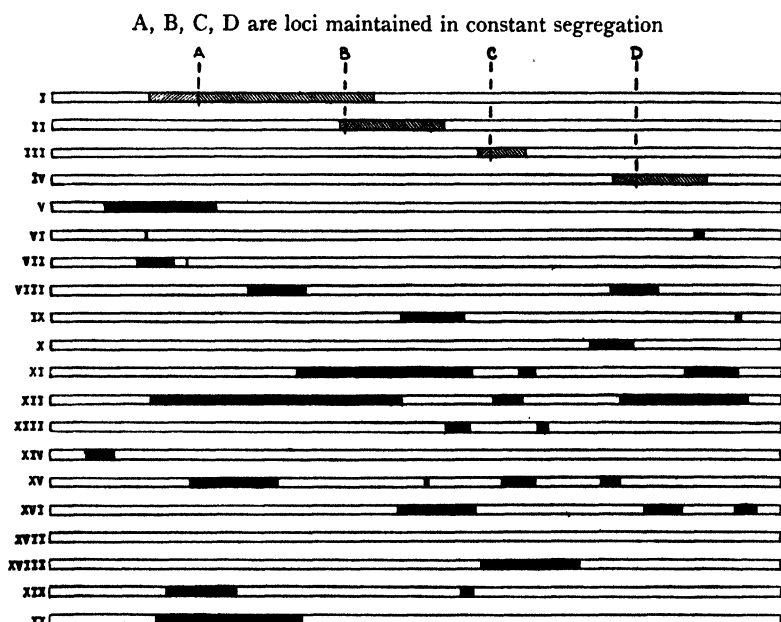
FIG. 4.—Model map of diploid organism with 20 chromosomes
after 10 generations of sib-mating.

In mice there are 20 pairs of chromosomes, with a total map length of about 2500 centimorgans. For various numbers of generations of sib-mating, the following table shows the expected numbers of tracts

of these two kinds, and the expected total map length occupied by them. The five diagrams give a representation, which is, however, not quantitatively accurate, of the speed and irregularity of the process by which heterogenic tracts are progressively diminished or eliminated (see Figs. 3-7). When the total number of tracts expected is m, the probability of there being none must be very nearly e^{-m}. We then see that after 30 generations this probability is not inappreciable, and must exceed $\frac{1}{2}$ before 40 generations.

A, B, C, D are loci maintained in constant segregation
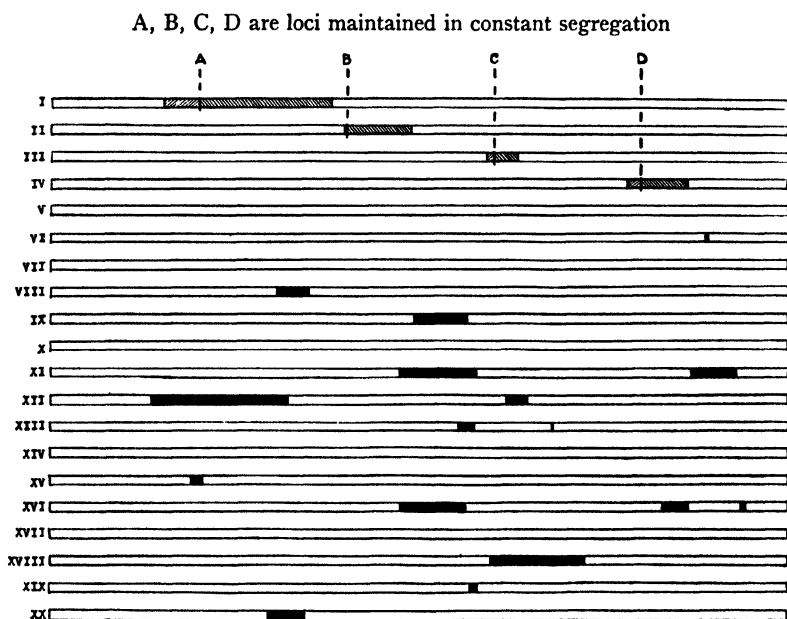
FIG. 5.—Model map of diploid organism with 20 chromosomes
after 15 generations of sib-mating.

The total number of junctions remaining when the whole material has become homogeneous is found by summing the rate of formation of junctions for all values of r, and dividing by 4, since each junction

E

formed has a probability of $\frac{1}{4}$ of surviving. This sums to

$$\frac{L}{20}\left\{\frac{4+24\epsilon}{1-\epsilon} + \frac{16-24\epsilon}{\frac{1}{2}+\epsilon}\right\} = 6L.$$

The average number of unbroken segments is therefore

$$6L+\gamma\nu.$$

A, B, C, D are loci maintained in constant segregation
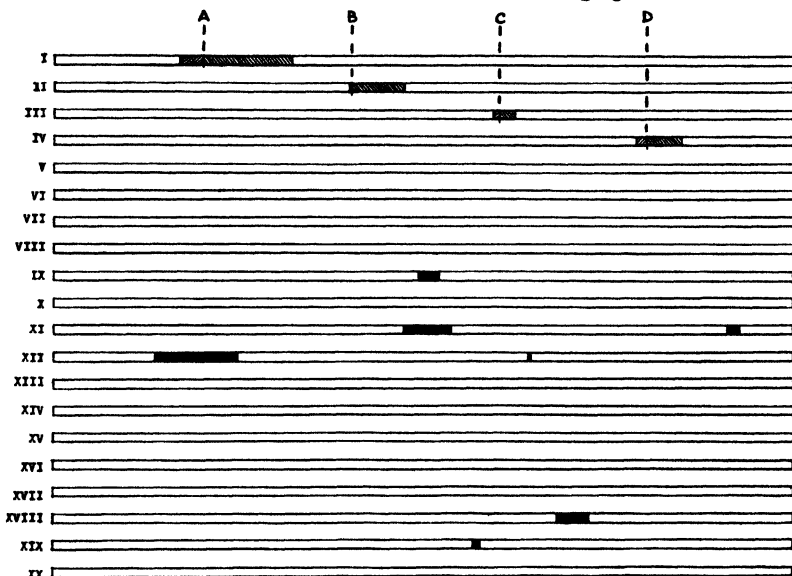
FIG. 6.—Model map of diploid organism with 20 chromosomes after 20 generations of sib-mating.

TABLE 11

Numbers and Expected Lengths of Heterogenic Tracts

Generations.	Expected Number.		Total.	Expected Total Length, Centimorgans.	Expected Length Attached to Each Dominant.
	End Tracts.	Enclosed Tracts.			
10	8·621	59·755	68·376	539·1	18
15	2·988	34·248	37·235	186·8	12·25
20	1·035	16·561	17·596	64·75	8·18
25	·359	7·366	7·724	21·87	6·43
30	·124	3·116	3·241	7·78	5·29

So with mice having 20 chromosome pairs and a total
map length about 2500 centimorgans, the average
number of unbroken segments in a line inbred by
sib-mating will be about 170.

Near the locus of a dominant gene carried
continuously heterozygous, the course of events is
comparatively simple. If the heterogeneous segment

A, B, C, D are loci maintained in constant segregation
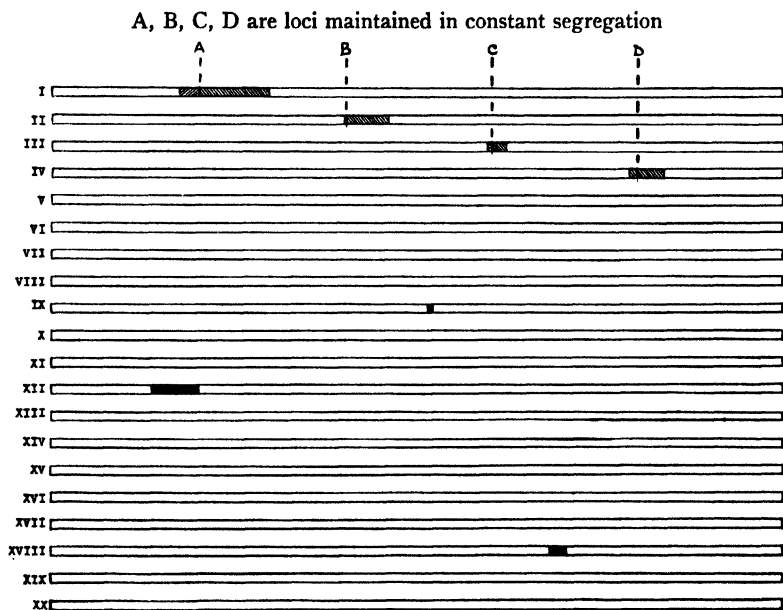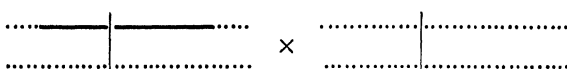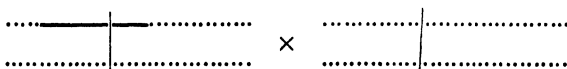
FIG. 7.—Model map of diploid organism with 20 chromosomes
after 25 generations of sib-mating.
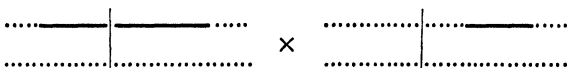
carried on one side of the dominant is of short length x,
then there is in each generation a small probability x
of a crossover in this region in the gamete carrying
the dominant, and in such a case the heterogeneous
segment is at once reduced in length, as explained
at the beginning of this section. There is, however,
an equal probability for the recessive gamete derived
from the heterozygous parent to have such a crossover ;

then the chance will be 1/3 for the region containing the junction to become homogeneous in the three recessive gametes of a later generation, while the chance is 2/3 that one of the recessive regions not containing the junction will do so. Consequently, the probability that the heterogeneous region will be diminished in length by means of crossing over in one or other of these gametes in any one generation is $4/3x$. It follows that the average length of the recessive segment after n generations will be approximately $3/4n$, or that the sum of the two segments carried on either side of a dominant gene will be $3/2n$ in length.
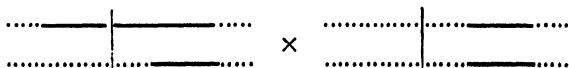
Original mating

Next generation (*a*)

following break in the strand carrying the dominant ;

or   (*b*)

leading in two cases out of three back to the original condition, and in the third case to

with the heterogenic region shortened as much as in (*a*), with a delay, however, averaging 6⅔ generations.

FIG. 8.—Modes of attrition of heterogenic region associated with a selected dominant.

It will be noticed that one quarter of these changes have a delayed effect, and it may be shown, by the method developed in the following chapter, that the average delay is 6⅔ generations. Consequently, the average delay including the crossovers which take effect

immediately is $1\frac{2}{3}$ generations. For this reason the values shown in the table are those of

$$\frac{3}{2(n-1\frac{2}{3})} = \frac{9}{6n-10}.$$

The course of events at a locus linked with one maintained permanently heterogenic is examined from another point of view in the following chapter (section 20).

IV

VARIOUS SYSTEMS OF INBREEDING

15. Other Generation Matrices

THIS chapter will be given to a number of inbreeding problems having generation matrices different from that of simple sib-mating, discussed in Chapter III. I do not know that any of these systems have been previously elucidated ; certainly, the practical results of examining them appear to be entirely unknown, and writers often discuss inbreeding as though the system of sib-mating were the only one available. Each of the special applications illustrated with the case of sib-matings in chapter III, may be developed with the various mating systems of the present chapter. In most cases the basis only of such applications is given in the form of the generation matrix or its principle components.

16. Approach to Homozygosity, Using Sib-matings, among Three Chromosomes, when the Fourth is Differentiated

If an inbred line is kept segregating for a factor D, by constantly inbreeding the phenotypes $D \times$ not-D, then of the gametes going to form each mating one always will carry D ; the three others will carry the same or different genes allelomorphic to D.

In mice, for example, a line could be kept segregating for yellow (the lethal gene A^Y), the three homologous loci being occupied by a, a^t, A or A^L. These are, of course, distinguishable, but if inbreeding were pursued subject only to the condition that each mating was to be a cross between yellow and non-yellow, the three chromosomes not containing the yellow gene would come in time by the process of inbreeding all to carry the same allelomorph.

When two dominants, sufficiently closely linked to obviate double crossing-over, are carried in the same chromosome, the same situation is produced in respect

of heterozygosity for any factors in the intervening tract, although these may not produce manifest effects. Thus if the normal allelomorphs of brown and misty are carried in the same chromosome, there will be about 8 per cent. crossovers, and if these are not used for mating, the included tract will remain differentiated from the corresponding tracts of the three doubly recessive chromosomes, which will, however, come gradually to be identical in gene content.

The differential segment of the sex chromosome is normally in this condition. In mammals the male in every mating is heterozygous for this tract, but the three X chromosomes involved in any mating will by inbreeding be made to be homozygous.

Representing the homologous locus or tract in the fourth chromosome by D, there are four types of sib-mating :—

Type.	Formula.	
w	$aa \times aD$	yielding only matings like itself
x	$ab \times aD$	yielding $(aa$ or $ab) \times (aD$ or $bD)$, so that out of 4 cases we obtain type w once, type x twice, and type y once
y	$aa \times bD$	yielding only $ab \times aD$, type x
z	$ab \times cD$	yielding $(ac$ or $bc) \times (aD$ or $bD)$, so that out of 4 cases we obtain type x twice and type z twice.

The generation matrix may therefore be written :-

<div align="center">TABLE 12</div>

		$\dfrac{w_0}{1}$	$\dfrac{x_0}{4}$	$\dfrac{y_0}{1}$	$\dfrac{z_0}{2}$
	w_1	1	1	.	.
$ab \times aD$	x_1	.	2	1	1
$aa \times bD$	y_1	.	1	.	.
$ab \times cD$	z_1	.	.	.	1

which may be read as four equations expressing w_1, x_1, y_1, and z_1 in terms of w_0, $\frac{1}{4}x_0$, y_0, and $\frac{1}{2}z_0$.

The mating types form a hierarchy, z involving three genes, x and y two, and w only one. From the diagonal block involving z we have the latent root

$$\lambda = \tfrac{1}{2}.$$

From that involving x and y the equation for λ is

$$-\lambda(2-4\lambda)-1 = 0$$

or

$$4\lambda^2-2\lambda-1 \quad = 0,$$

with roots

$$\lambda = \epsilon, \qquad \tfrac{1}{2}-\epsilon,$$

where, as before, ϵ is written for

$$\tfrac{1}{4}(1+\sqrt{5}).$$

There are only three latent roots in place of six for the remainder of the germ-plasm, but the dominant root has the same numerical value.

The principal components corresponding with these roots are then easily found to be

Root.	Linear Function of the Frequencies.	Designation.
ϵ	$x+(4\epsilon-2)y+2\epsilon z$	L
$\tfrac{1}{2}-\epsilon$	$x-4\epsilon y-(2\epsilon-1)z$	M
$\tfrac{1}{2}$	z	z

Expressing x, y, and z explicitly in terms of the linear functions L, M, and z, we have

$$10x = 4(\epsilon+1)L+(6-4\epsilon)M-10z$$
$$10y = (4\epsilon-1)L-(4\epsilon-1)M- 5z$$

whence, after n generations, neglecting all but the dominant root,

$$10(x_n+y_n) = (8\epsilon+3)\epsilon^n L_0$$
$$= \epsilon^n\{(8\epsilon+3)x_0+(12\epsilon+2)y_0+(14\epsilon+4)z_0\}.$$

The complexities of the mating types x, y, and z are therefore

				Mating Type.	Complexity.	Numerical.
x	.	.	.	$ab \times aD$	$(8\epsilon+3)/10$	·94721
y	.	.	.	$aa \times bD$	$(12\epsilon+2)/10$	1·17082
z	.	.	.	$ab \times cD$	$(14\epsilon+4)/10$	1·53262

The proportion of x to y tends in the limit to the frequencies

				Mating Type.	Frequency.	Numerical per cent.
x	.	.	.	$ab \times aD$	$4-4\epsilon$	76·3932
y	.	.	.	$aa \times bD$	$4\epsilon-3$	23·6068
					1	100·0000

17. Fate of Particular Genes, or Junctions

In some cases it is useful to distinguish matings not only by their type, but further by their genic content. For example, in following the consequences of sib-mating in the neighbourhood of the locus of a factor kept segregating, we may use the generation matrix

TABLE 13

			$w/1$	$x/4$	$y/1$	$y'/1$	$x'/4$	$w'/1$
$aa \times aD$	.	. w	1	1	.	.	.	.
$ab \times aD$	.	. x	.	1	1	.	1	.
$aa \times bD$	.	. y	.	1	.	.	.	.
$bb \times aD$	.	. y'	.	.	.	.	1	.
$ab \times bD$	.	. x'	.	1	.	1	1	.
$bb \times bD$	.	. w'	.	.	.	.	1	1

in which each of the mating types previously distinguished has been divided into two, mutually convertible by the substitution (ab).

Taking the matrix as a whole, including the first and last rows, the determinant of the λ matrix is

$$(1-\lambda)^2(1-4\lambda^2)(1+2\lambda-4\lambda^2)$$

giving the six latent roots

$$+\tfrac{1}{2}, \ -\tfrac{1}{2}, \ \epsilon, \ \epsilon-\tfrac{1}{2} \text{ and } 1 \text{ twice.}$$

Corresponding with the two unit roots, there are two principal components

$$
\begin{array}{ll}
L & 3w+2x+2y+\ y'+\ x' \\
L' & x+\ y+2y'+2x'+3w'.
\end{array}
$$

For the remaining rational roots we find,

$$
\begin{array}{lll}
\lambda & & \\
\tfrac{1}{2} & x+2y-x'-2y' & M, \\
-\tfrac{1}{2} & x-2y-x'+2y' & N,
\end{array}
$$

while the two irrational roots give,

$$
\begin{array}{lll}
\lambda & & \\
\epsilon & \epsilon\,x+y+\ \ \epsilon\,x'+y' & P, \\
\tfrac{1}{2}-\epsilon & (\tfrac{1}{2}-\epsilon)x+y+(\tfrac{1}{2}-\epsilon)x'+y' & Q.
\end{array}
$$

To find the frequency of any mating type, such as w', after n generations of sib-mating, we express w' in terms of the principal components

$$
w' = \tfrac{1}{3}L'+\tfrac{1}{8}M+\tfrac{1}{24}N-\tfrac{1}{10}(1+6\epsilon)P+\tfrac{1}{10}(6\epsilon-4)Q
$$

so that if the initial mating is of type $ab \times aD$, giving $x_0 = 1$, it follows that

$$
\begin{aligned}
w'_n &= \tfrac{1}{3}L'_0+\tfrac{1}{8}(\tfrac{1}{2})^n M_0+\tfrac{1}{24}(-\tfrac{1}{2})^n N_0-\tfrac{1}{10}(1+6\epsilon)\epsilon^n P_0+ \\
&\qquad\qquad\qquad\qquad \tfrac{1}{10}(6\epsilon-4)(\tfrac{1}{2}-\epsilon)^n Q_0 \\
&= \tfrac{1}{3}+\tfrac{1}{8}(\tfrac{1}{2})^n+\tfrac{1}{24}(-\tfrac{1}{2})^n-\tfrac{1}{20}(3+8\epsilon)\epsilon^n-\tfrac{1}{20}(7-8\epsilon)(\tfrac{1}{2}-\epsilon)^n.
\end{aligned}
$$

The limiting value is $\tfrac{1}{3}$, as was to be expected. To find the average number of generations required in those cases in which the mating of type w' is ultimately established, we require

$$
3\sum_{n=1}^{\infty}(w'_n-w'_{n-1})n
$$

$$
= \sum_{n=0}^{\infty}(1-3w'_n)
$$

$$
= -\frac{3}{8(\tfrac{1}{2})} - \frac{1}{8}\cdot\frac{1}{1+\tfrac{1}{2}} + \frac{3}{20}\cdot\frac{3+8\epsilon}{1-\epsilon} + \frac{3}{20}\cdot\frac{7-8\epsilon}{\tfrac{1}{2}+\epsilon}
$$

$$
= -\tfrac{3}{4} - \tfrac{1}{12} + \tfrac{3}{20}(14+44\epsilon) + \tfrac{3}{20}(36-44\epsilon)
$$

$$
= -\tfrac{5}{6}+7\tfrac{1}{2} = 6\tfrac{2}{3}.
$$

Consequently there is an average delay of $6\frac{3}{5}$ generations in the establishment of the homogeneous mating $bb \times bD$ in this $\frac{1}{5}$ of cases, while the $\frac{4}{5}$ which end in $aa \times aD$ are completed more quickly. The actual distribution of the delay can of course be obtained from w'_n.

18. Parent-offspring Inbreeding

Inbreeding may be carried out progressively by using parent-offspring matings, in which each individual is used in two successive generations, first as offspring and next as parent. The system differs from that of sib-mating, first in that matings between individuals having no gene in common cannot occur (or at least can only occur after an irregularity such as a sib-mating introduced in the inbred line), secondly in that the mating between homozygote and heterozygote (type u) must now be subdivided into two mating-types, according as the parent or the offspring is homozygous.

Following out the genetic consequences of these matings, we find the following λ-matrix :—

<div align="center">TABLE 14</div>

	Offspring.	Parent.	$\dfrac{u_0}{2}$	$\dfrac{u'_0}{2}$	$\dfrac{v_0}{2}$	$\dfrac{x_0}{4}$	$\dfrac{w_0}{1}$
u_1 . . .	aa	ab	-2λ	1	1	1	.
u'_1 . . .	ab	aa	1	-2λ	.	.	1
v_1 . . .	ab	ab	.	1	$1-2\lambda$	1	.
x_1 . . .	ab	ac	.	.	.	$2-4\lambda$	.
w_1 . . .	aa	bb	.	.	.	.	$-\lambda$

The mating type w has been included in order to illustrate the evaluation of its complexity when it arises, as it may, through an irregularity in the mating-system.

The latent roots and corresponding principal components are :—

$\frac{1}{2}$	x
0	$u-v$
ϵ	$2u+4\epsilon u'+4\epsilon v+(4\epsilon+1)x+4w$
$\frac{1}{2}-\epsilon$	$2u+(2-4\epsilon)u'+(2-4\epsilon)v+(3-4\epsilon)x+4w.$

For the largest root, ϵ, the frequencies of the three surviving mating-types are :—

Offspring.		Parent.			Per Cent.
aa	$\times$	ab	u	$2-2\epsilon$	38·1966
ab	$\times$	aa	u'	$-3+4\epsilon$	23·6068
ab	$\times$	ab	v	$2-2\epsilon$	38·1966

and the complexities, in absolute measure :—

Offspring.		Parent.		$\div 5.$	
aa	$\times$	ab	U	$2(1+\epsilon)$	·7236
ab	$\times$	aa	U'	$(1+6\epsilon)$	1·1708
ab	$\times$	ab	V	$(1+6\epsilon)$	1·1708
ab	$\times$	ac	X	$(2+7\epsilon)$	1·5326
aa	$\times$	bb	W	$4(1+\epsilon)$	1·4472.

It is remarkable that this system should give a rate of progress towards inbreeding exactly equal to that of successive sib-matings. It is still more remarkable that the intercalation of a single sib-mating in such a series should retard progress to exactly the same extent as the intercalation of a parent-offspring mating in a series of sib-matings.

For a sib-mating intercalated in a series of parent-offspring matings, which will be resumed by the use of one of the sibs as parent in the next generation, we find the following matrix of probabilities :—

TABLE 15

Last Parent-offspring Mating.

				u	u'	v
Sib-mating u .	.	.		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
,,	u'	.	.	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
,,	v .	.	.	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
,,	w	.	.	.	.	$\frac{1}{8}$

It will be noted that, following a last parent-offspring mating of type v, which will produce two unlike types of homozygote, it is possible that the following sib-mating shall be of type w. Multiplying the columns

by the frequencies of u, u', and v determined above, we find :—

$$\tfrac{1}{4}U + \tfrac{1}{4}U' + \tfrac{1}{4}V + \tfrac{1}{4}(1-\epsilon)W,$$

and inserting the values for U, V, etc., we have :—

	÷20
$\tfrac{1}{4}U$	$2 + 2\epsilon$
$\tfrac{1}{4}U'$	$1 + 6\epsilon$
$\tfrac{1}{4}V$	$1 + 6\epsilon$
$\tfrac{1}{4}(1-\epsilon)W$	$3 - 2\epsilon$
	$7 + 12\epsilon$ or ·8354102,

identically the same result as that obtained for a parent-offspring mating intercalated in a sib-mating series.

19. Effect of Parent-offspring Inbreeding on the Differential X-Segment or on any Tract Carried in Constant Segregation

With the designations shown in the table we find the following λ-matrix.

TABLE 16

	Offspring.	Parent.	u_0	u'_0	v_0	v'_0
u_1	ab	aD	$-\lambda$	$\tfrac{1}{2}$	.	1
u'_1	aD	ab	1	$-\lambda$	.	.
v_1	aa	bD	.	.	$-\lambda$	.
v'_1	bD	aa	.	.	.	λ

Only in u and u' have the mates a gene in common, as they must have if produced by parent-offspring mating. The equation for the latent root is

$$\lambda = 0, \quad \text{or} \quad \lambda^2 = \tfrac{1}{2},$$

giving two roots of equal absolute magnitude. Progress towards homozygosity is thus considerably (1·635 times) more rapid than in other parts of the germ-plasm with parent-offspring matings, and than any part of the germ-plasm with sib-matings, being half as rapid as selfing diploids. Numerically this is ·34658 per generation.

The principal components corresponding to the two roots are

$$L = u + \tfrac{1}{2}u'\sqrt{2} + \tfrac{1}{2}v'\sqrt{2},$$
$$M = u - \tfrac{1}{2}u'\sqrt{2} - \tfrac{1}{2}v'\sqrt{2}.$$

Hence the expressions for u and u' in terms of L, M, and v' are

$$u = \tfrac{1}{2}L + \tfrac{1}{2}M$$
$$u' = \frac{1}{\sqrt{2}}L - \frac{1}{\sqrt{2}}M - 2v'.$$

Then, after n generations

$$u_n = \tfrac{1}{2}L_n + \tfrac{1}{2}M_n$$
$$= \frac{1}{2}\left\{\left(\frac{1}{\sqrt{2}}\right)^n L_0 + \left(-\frac{1}{\sqrt{2}}\right)^n M_0\right\}.$$
$$= 2^{-\frac{1}{2}n}u_0 \text{ when } n \text{ is even}$$
$$\text{but} \quad 2^{-\frac{1}{2}n}(\tfrac{1}{2}u'_0\sqrt{2} + v'_0\sqrt{2}) \text{ if } n \text{ is odd.}$$

Similarly

$$u'_n = 2^{-\frac{1}{2}n}(\tfrac{1}{2}u'_0\sqrt{2} + v'_0\sqrt{2}) \text{ if } n \text{ is even}$$
$$\text{but} \quad 2^{-\frac{1}{2}n}u_0 \text{ if } n \text{ is odd.}$$

Since no root exceeds all others in absolute value there is no limiting frequency distribution.

In this case it is of interest to set up a cycle of sib and parental matings, which may be done in a variety of ways. For example, in three generations we may have two parental and one sib mating, in accordance with the pedigree :—
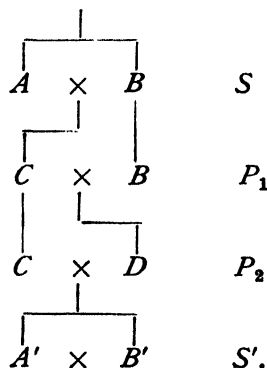
Such a scheme can be carried out in two ways, according as A (and therefore C) is dominant or recessive. Let us suppose that A is dominant (or male). Then the two possible mating types for S, x, and y will produce the two possible mating types u, u' in P_1 according to the table.

TABLE 17

Generation Matrix for Sib-mating

			$ab \times aD$	$aa \times bD$
Offspring.	Parent.		x	y
ab	aD	u	.	.
aD	ab	u'	1	.

Next, the two possible mating types in P_1 will yield the same in P_2 according to the second matrix :—

TABLE 18

Generation Matrix for First Parent Mating

	u	u'
u .	.	$\frac{1}{2}$
u' .	1	.

Finally, the two possible types of sib-mating will be produced by these according to the matrix :—

TABLE 19

Generation Matrix for Second Parent Mating

	u	u'
x .	$\frac{1}{2}$	$\frac{1}{2}$
y .	$\frac{1}{4}$	$\frac{1}{4}$

Multiplying the three matrices together we obtain for three generations :—

TABLE 20

Matrix for Cycle of Three Generations

	x	y
x .	$\frac{1}{4}$	.
y .	$\frac{1}{8}$	.

The latent roots are 0 and $\frac{1}{4}$; as this is the result of three generations we may put

$$\lambda^3 = \tfrac{1}{4}$$
$$\lambda = \cdot 6299605249$$
$$-\log_e\lambda = \cdot 462098120$$

as a measure of the rate of progress on this system in the parts of the germ-plasm concerned. It exceeds even that for parent-offspring matings ($\cdot 34658$), and is more than double as great as the speed of sib-matings.

Such a scheme is capable, therefore, of greatly accelerating inbreeding in a marked portion of the germ-plasm. It will, however, only be advantageous if it does not greatly retard progress in portions inherited independently. For these we find the three matrices :—

<div align="center">TABLE 21</div>

		u	v	w	S			u	u'	v	P_1			u	u'	v	P_2
	u	$\frac{1}{4}$	$\frac{1}{2}$	.			u	.	$\frac{1}{2}$	$\frac{1}{2}$			u	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
P_1	u'	$\frac{1}{4}$	.	1		P_2	u'	$\frac{1}{2}$	.	.		S'	v	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
	v	$\frac{1}{4}$	$\frac{1}{2}$	.			v	.	$\frac{1}{2}$	$\frac{1}{2}$			w	.	.	$\frac{1}{4}$

of which the product is :—

<div align="center">TABLE 22</div>

		$\dfrac{u}{32}$	$\dfrac{v}{32}$	$\dfrac{w}{16}$	S
	u	10	12	8	
S'	v	5	6	4	
	w	1	1	1	

The latent roots are 0 and $\frac{1}{32}(9 \pm \sqrt{73})$.
The dominant root is therefore

$$\lambda^3 = \cdot 548250117$$
$$\lambda = \cdot 818451428$$
$$-\log_e\lambda = \cdot 200341226,$$

showing that progress on the whole is retarded by about 5·47 per cent.

If, therefore, it is thought that any characters of importance, for example in dairy cattle, are carried in the sex chromosome, it is possible relatively rapidly to render this region homozygous, without much retarding the progress towards homozygosity in the autosomes.

The reader should try the effect of using a recessive or female as animal A ; a second variation when A is a dominant is to use it a third time. It would be interesting also to trace the effect on genes linked with the marker, as in the following section.

20. Loci Linked with Heterogenic Loci

The condition at any locus in the same chromosome as a segregating dominant may be specified, writing first that one of the four genes which is in the same chromosome as the dominant gene, as one of ten possible conditions, as follows :—

<div align="center">TABLE 23</div>

1. $ab \times bb$	6. $ab \times ac$
2. $ba \times bb$	7. $ba \times ac$
3. $bb \times ab$	8. $aa \times bc$
4. $ab \times ab$	9. $bc \times aa$
5. $aa \times bb$	10. $ab \times cd$

Then, if y stands for the recombination fraction with the locus of the dominant, the λ-matrix for a single generation of sib-mating is shown in Table 24.

The 5×5 determinant for the digenic cases is reducible to the quintic expression

$$(1-2y)(1-2y+2y^2)^2$$
$$+\lambda \ (1-2y)^3(1-2y+2y^2)$$
$$-\lambda^2(10-38y+68y^2-76y^3+48y^4-16y^5)$$
$$+4y\lambda^3$$
$$+\lambda^4(24-32y+16y^2)$$
$$-16\lambda^5.$$

<div align="right">F</div>

TABLE 24

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.	$(1-y)^2-\lambda$	y^2	·	$1-2y+2y^2$	·	y^2	$(1-y)^2$	·	·	·
2.	·	$-\lambda$	1	$2y-2y^2$	·	·	·	·	·	·
3.	y^2	$(1-y)^2$	$1-4\lambda$	1	·	$1-y^2$	$2y-2y^2$	·	·	·
4.	$y(1-y)$	$y(1-y)$	1	$1-4\lambda$	1	$3y-3y^2$	$3y-3y^2$	1	$2y-2y^2$	$2y(1-y)$
5.	·	·	·	$1-2y+2y^2$	$-\lambda$	·	·	·	·	·
6.						$1-y^2-4\lambda$	$2y-y^2$	1	·	$2y(1-y)$
7.						$1-2y+3y^2$	$2-4y+3y^2-4\lambda$	·	$1-2y+2y^2$	$1-2y+2y^2$
8.						$1-2y+y^2$	y^2	-2λ	·	·
9.						y^2	$(1-y)^2$	·	$-\lambda$	·
10.						·	·	·	·	$1-2y+2y^2-2\lambda$

This has one linear factor,

$$1 - 2y + 2y^2 - 2\lambda,$$

which leaves, for the remaining four roots, the equation,

$$
\begin{aligned}
& (1-2y)(1-2y+2y^2) \\
& +\lambda\,(1-2y)(3-4y+4y^2) \\
& -2\lambda^2(2-5y+8y^2-4y^3) \\
& -8\lambda^3(1-y) \\
& +8\lambda^4 = 0.
\end{aligned}
$$

In terms of the recombination fraction y, the dominant root takes the values of the following table :—

TABLE 25

Speed in Terms of Recombination Fraction

y per cent.	λ.	$-\log_e \lambda$.	Speed per cent. of Unlinked Loci.
0	1·000000	0	0
5	·942277	·05945	28·05
10	·898358	·10719	50·58
15	·870419	·13879	65·49
20	·8513774	·16090	75·92
25	·8374848	·17735	83·67

The 4×4 determinant also for the trigenic cases has a single linear factor, namely

$$1 - 2y + 2y^2 + 4\lambda$$

corresponding to the root $-\frac{1}{4}$ when y is $\frac{1}{2}$, and to the root $-\frac{1}{4}$ when y is zero. The remaining factor is

$$
\begin{aligned}
& (1-2y)(1-2y+2y^2) \\
& -4\lambda^2(2-3y+2y^2) \\
& +8\lambda^3.
\end{aligned}
$$

When y is zero, the roots are those already found for the recessive allelomorphs of a dominant factor kept constantly segregating.

As in the case of an unlinked factor the root for the tetragenic mating is the same as one of those of the digenic matings, namely

$$\tfrac{1}{2}(1 - 2y + 2y^2).$$

Corresponding with any latent root we can, as always, obtain the principal component and the frequency distribution. Thus for the case of 5 per cent. recombination with $\lambda = \cdot 942277$, the limiting frequency distribution for cases not yet fully homogenic will be

TABLE 26

			5 per cent.	Limit, $y = 0$. Frequency.	Complexity.
1. $ab \times bb$	.	.	73·378	$3/y$	3
2. $ba \times bb$	.	.	2·491	1	1
3. $bb \times ab$	.	.	8·167	4	1
4. $ab \times ab$	.	.	12·873	8	2
5. $aa \times bb$	.	.	3·091	2	2

Thus at a recombination distance of 5 per cent. of 100 cases not yet homogenic nearly three quarters will be composed as if no crossover had yet occurred in the intervening segment, while in one quarter the gene in the chromosome carrying the dominant is also present in at least one, but not in all, of the chromosomes bearing recessives.

With alternate matings of parent and offspring, we must specify not only which mate is parent and which offspring, but also which of the four gametes represented carries the dominant marker, or the Y of the sex chromosome (these are shown by bold type). Of the mating types possible only ten can arise from matings of parent and offspring. These give the λ-matrix, in Table 27.

For the three trigenic cases, the determinant has the factors
$$\lambda$$
$$1 - 2y + 2y^2 - 2\lambda^2$$

giving a zero and two equal roots of opposite sign. While the seven digenic cases give the equations

$$\lambda^3 = 0$$

or

$$\lambda^4 - \tfrac{1}{2}\lambda^2(3 - 4y + 2y^2) + \tfrac{1}{4}(2 - 5y + 4y^2 - 2y^3) = 0.$$

TABLE 27

Linked Locus with Parental Matings

| P O	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. $ab \times bb$	$-\lambda$	$1-y$	.	.	.	.	$\frac{1}{2}(1-y)$	.	.	$\frac{1}{2}(1-y)$
2. $bb \times ab$	$1-y$	$-\lambda$	y	.	.	.	.	.	.	.
3. $ba \times bb$	.	.	$-\lambda$	.	.	.	$\frac{1}{2}y$	.	.	.
4. $aa \times ab$	.	.	.	$-\lambda$	$\frac{1}{2}$	.	.	.	.	.
5. $ab \times aa$	.	.	.	$\frac{1}{2}$	$-\lambda$	$\frac{1}{2}$	.	$\frac{1}{2}(1-y)$	$\frac{1}{2}y$	.
6. $ab \times ab$	.	y	.	.	.	$-\lambda$	$\frac{1}{2}$	.	.	$\frac{1}{2}y$
7. $ab \times ab$	.	.	.	$\frac{1}{2}$	.	$\frac{1}{2}$	$-\lambda$	$\frac{1}{2}(1-y)$	$\frac{1}{2}y$	.
8. $ab \times ac$	.	.	.	.	.	.	.	$-\lambda$	.	$\frac{1}{2}y$
9. $ab \times ac$	.	.	.	.	.	.	.	.	$-\lambda$	$\frac{1}{2}(1-y)$
10. $ab \times ac$	.	.	.	.	.	.	.	y	$1-y$	$-\lambda$
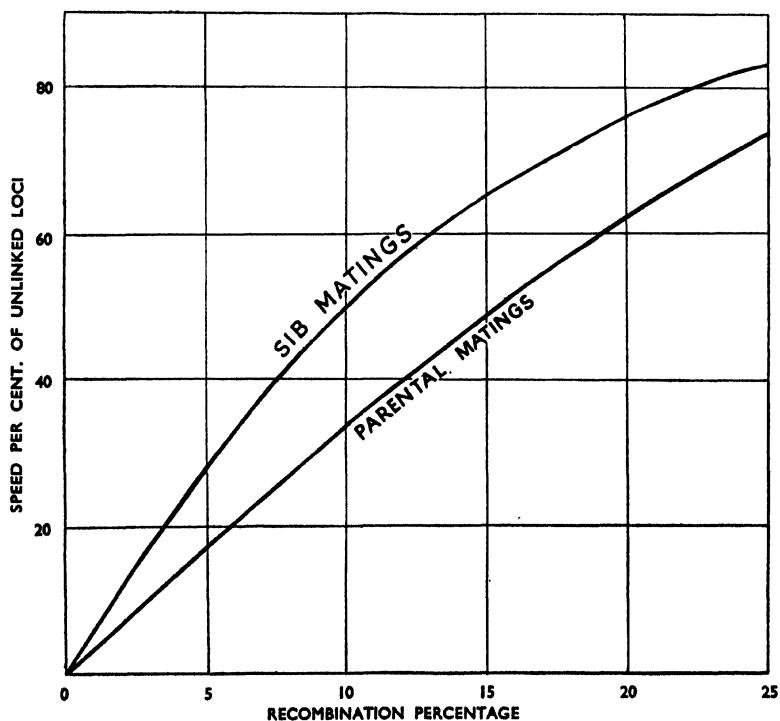
FIG. 9.—The relation between speed of approach towards homozygosity and linkage with a locus constantly segregating.

TABLE 28

Speed in Terms of Recombination Fraction

y per cent.	λ.	$\log_e 1/\lambda$.	Speed per cent. of Unlinked Loci.
0	1·0000000	0	0
5	·9637827	·03689	17·41
10	·9308963	·07161	33·79
15	·9011291	·10411	49·13
20	·8759171	·13248	62·51
25	·8551887	·15643	73·81

The reduction in the length of the segments attached to a dominant is slower than with sib-matings, chiefly because a new chromosome including the dominant is only introduced once in two generations. This is only partially counterbalanced by the greater efficiency of crossover segments bearing the corresponding recessive.

21. Polysomic Segregation

Polysomic organisms differ from disomic in having more than two chromosomes mutually homologous and capable of pairing, and of interchange of segments. Organisms with sets of four homologous are tetrasomic, with six hexasomic. The gametes of tetrasomic forms are diploid, those of hexasomic forms are triploid.

For any chosen factor disomic organisms are classified as homozygous or heterozygous according to whether the two allelomorphs are equivalent or unlike. For polysomic organisms the parallel classification is more complex. With disomic organisms there are only two alternatives because there are only two partitions of the number 2, namely those written (2) for the homozygote, and (1^2) for the heterozygote. There are five partitions of the number 4, and eleven of the number 6. These are set out below, with representative gametic formulæ. These last may be written in full, or contracted on the same system as chemical formulæ.

TABLE 29

Partitions of Four.	Genetical Formula.	Contraction.	Descriptive Term.	
(4)	*aaaa*	a_4	homogenic	nulliplex or quadruplex
(31)	*aaab*	a_3b	heterogenic-(31)	simplex or triplex
(2²)	*aabb*	a_2b_2	heterogenic-(2²)	duplex
(21²)	*aabc*	a_2bc	heterogenic-(21²)	trigenic
(1⁴)	*abcd*	*abcd*	heterogenic-(1⁴)	tetragenic

Partitions of Six.				
(6)	...	a_6	homogenic	nulliplex or sextuplex
(51)	...	a_5b	heterogenic-(51)	simplex or quintuplex
(42)	...	a_4b_2	heterogenic-(42)	duplex or quadruplex
(3²)	...	a_3b_3	heterogenic-(3²)	triplex
(41²)	...	a_4bc	heterogenic-(41²)	trigenic
(321)	...	a_3b_2c	heterogenic-(321)	trigenic
(2³)	...	$a_2b_2c_2$	heterogenic-(2³)	trigenic
(31³)	...	a_3bcd	heterogenic-(31³)	tetragenic
(2²1²)	...	a_2b_2cd	heterogenic-(2²1²)	tetragenic
(21⁴)	...	a_2bcde	heterogenic-(21⁴)	pentagenic
(1⁶)	...	*abcdef*	heterogenic-(1⁶)	hexagenic

The descriptive terms hitherto available, based on the number of dominant genes, go only a short way towards drawing the distinctions necessary.

The frequency of the different possible types of gamete which any heterogenic organism can produce may be derived by the same method as Fisher and Mather (6) have used for simplex, duplex and triplex organisms ; this is that with tetrasomic organisms every selection of two out of the four available occurs with equal frequency, these making together the fraction $1 - a$ of all gametes, while the remaining gametes containing two twin loci are equally divided among the four ways in which these may be selected. With hexasomic organisms, also, all the twenty selections of three out of the six available occur with equal frequency, the total of these being $1 - \beta$, while the remainder, β, is divided equally among the thirty ways of using a pair of twin loci and one other.

The gametic output of the four types of heterogenic tetrasomics is therefore :—

TABLE 30

Parent.	a_2	ab	b_2
a_2b a_2b_2	$\frac{1}{4}(2+\ a)$ $\frac{1}{6}(1+2a)$	$\frac{1}{4}(2-2a)$ $\frac{1}{6}(4-4a)$	$\frac{1}{4}a$ $\frac{1}{6}(1+2a)$

The trigenic genotype a_2bc gives six types of gamete as follows

$$\tfrac{1}{12}(2+4a)a_2,\ \tfrac{1}{12}(4-4a)ab,\ \tfrac{1}{12}(4-4a)ac,$$
$$\tfrac{1}{12}(3a)b_2,\ \tfrac{1}{12}(3a)c_2,\ \tfrac{1}{12}(2-2a)bc.$$

The tetragenic genotype $abcd$ gives

$$\tfrac{1}{4}a \text{ each of } a_2,\ b_2,\ c_2,\ d_2\ ;$$
and $\tfrac{1}{6}(1-a)$ each of $ab,\ ac,\ ad,\ bc,\ bd,\ cd$.

The gametes from the three kinds of digenic hexasomics are as given by Fisher and Mather :—

TABLE 31

Parent.	a_3	a_2b	ab_2	b_3	
a_3b	$3+2\beta$	$3-3\beta$	β	.	$\div 6$
a_2b_2	$3+3\beta$	$9-5\beta$	$3+\beta$	β	$\div 15$
a_3b_3	$1+3\beta$	$9-3\beta$	$9-3\beta$	$1+3\beta$	$\div 20$

Similarly, the trigenic hexasomics give :—

TABLE 32

Parent.	a_3	b_3	c_3	a_2b	ab_2	b_2c	bc_2	ac_2	a_2c	abc	
a_4bc	$6+6\beta$	.	.	$9-5\beta$	4β	β	β	4β	$9-5\beta$	$6-6\beta$	$\div 30$
a_2b_2c	$3+9\beta$	4β	.	$18-6\beta$	$9+3\beta$	$3+\beta$	4β	6β	$9-3\beta$	$18-18\beta$	$\div 60$
$a_2b_2c_2$	2β	2β	2β	$3+\ \beta$	$3+\ \beta$	$3+\beta$	$3+\beta$	$3+\beta$	$3+\ \beta$	$12-12\beta$	$\div 30$

With four genes there will be twenty different possible kinds of gametes. Thus a parent of type a_3bcd out of sixty gametes will give :—

TABLE 33

$3+9\beta$	of type . .	a_3	$3+ 9\beta$
$9-3\beta$	each of types	a_2b, a_2c, a_2d	$27- 9\beta$
6β	each of types	ab_2, ac_2, ad_2	$+18\beta$
$9-9\beta$	each of types	abc, abd, acd	$27-27\beta$
2β	each of types	$b_2c, bc_2, b_2d, bd_2, c_2d, cd_2$ . .	$+12\beta$
$3-3\beta$	of type . .	bcd	$3- 3\beta$

60

While the genotype a_2b_2cd yields out of sixty :—

TABLE 34

4β	each of types	a_3, b_3	$+ 8\beta$
$6+ 2\beta$	each of types	a_2b, ab_2	$12+ 4\beta$
$3+ \beta$	each of types	a_2c, a_2d, b_2c, b_2d . . .	$12+ 4\beta$
4β	each of types	ac_2, ad_2, bc_2, bd_2 . . .	$+16\beta$
2β	each of types	c_2d, cd_2	$+ 4\beta$
$12-12\beta$	each of types	abc, abd	$24-24\beta$
$6- 6\beta$	each of types	acd, bcd	$12-12\beta$

60

A pentagenic genotype has thirty-five possible kinds of gamete ; thus a_2bcde gives out of sixty :—

TABLE 35

4β	of type . .	a_3	$+ 4\beta$
$3+ \beta$	each of types	a_2b, a_2c, a_2d, a_2e . . .	$12+ 4\beta$
4β	each of types	ab_2, ac_2, ad_2, ae_2 . . .	$+16\beta$
$6-6\beta$	each of types	$abc, abd, abe, acd, ace, ade$.	$36-36\beta$
2β	each of types	$b_2c, b_2d, b_2e, bc_2, c_2d, c_2e, bd_2, cd_2, d_2e,$	
		be_2, ce_2, de_2	24β
$3-3\beta$	each of types	bcd, bce, bde, cde . . .	$12-12\beta$

60

Finally, the hexagenic parent $abcdef$ gives twenty kinds of gamete like abc, each with frequency $(1-\beta)/20$, and thirty kinds of gamete like a_2b each with frequency $\beta/30$.

22. Selfing Tetrasomic Organisms

On self-fertilisation a polysomic organism of given partitional type will reproduce its own and other types with frequencies expressible as quadratic functions of a or β. Thus the tetrasomic genotype digenic-(31), with gametic output,

$$\begin{array}{cccc} \text{Gamete} & a_2 & ab & b_2 \\ \text{Frequency} & 2+a & 2-2a & a & \div 4, \end{array}$$

will on self-fertilisation produce homogenic offspring with frequency

$$(2+a)^2+a^2 \qquad\qquad = 4+4a+2a^2 \quad \div 16$$

heterogenic-(31) offspring with frequency

$$2(2+a)(2-2a)+2(2-2a)a = 8 \qquad -8a^2 \quad \div 16$$

and heterogenic-(2^2) offspring with frequency

$$2(2+a)a+(2-2a)^2 \qquad = 4-4a+6a^2 \quad \div 16.$$

The generation matrix for such plants will in general be of the second degree in a in each element. Thus for selfing tetrasomics we have, omitting the frequencies homogenic :—

TABLE 36

Generation Matrix for Selfing a Tetrasomic Organism

Parent.

		$\dfrac{f_0(31)}{16}$	$\dfrac{f_0(2^2)}{36}$	$\dfrac{f_0(21^2)}{72}$	$\dfrac{f_0(1^4)}{12}$
	$f_1(31)$	$8 \qquad -8a^2$	$16+16a-32a^2$	$16+52a-68a^2$	$12a-12a^2$
	$f_1(2^2)$	$4-4a+6a^2$	$18-24a+24a^2$	$18-24a+51a^2$	$2-4a+11a^2$
Offspring.	$f_1(21^2)$	.	.	$36-36a$	$8-4a-4a^2$
	$f_1(1^4)$	.	.	.	$2-4a+2a^2$

Since the number of different genes cannot increase, the matrix is hierarchical. The diagonal block for the tetragenic genotype yields the latent root

$$2-4a+2a^2 = 12\lambda, \qquad \lambda = \tfrac{1}{6}(1-a)^2,$$

with the corresponding principal component

$$f(1^4),$$

so that

$$f_n(1^4) = \{\tfrac{1}{6}(1-a)^2\}^n f_0(1^4).$$

That for the trigenic genotype likewise gives

$$36-36a = 72\lambda, \qquad \lambda = \tfrac{1}{2}(1-a),$$

with the principal component

$$f(21^2)+2f(1^4).$$

The diagonal block for the digenic genotypes yields the quadratic equation

$$(8-8a^2-16\lambda)(18-24a+24a^2-36\lambda) =$$
$$(16+16a-32a^2)(4-4a+6a^2)$$

or

$$\lambda^2-\tfrac{1}{6}(6-4a+a^2)\lambda+\tfrac{1}{36}(5-12a+9a^2-2a^3) = 0,$$

giving the two solutions

$$\tfrac{1}{6}(5-2a) \quad \text{and} \quad \tfrac{1}{6}(1-a)^2.$$

The four latent roots with the corresponding principal components are shown below. For all values of a the order of the roots is unaltered, and two are always equal. It will be remembered that any linear function of the two given for the double root might equally be regarded as a principal linear component.

The principal linear components of frequency corresponding with the root $(1-a)^2/6$ are

$$f(1^4) \text{ and}$$

$$3(2-2a+3a^2)f(31)-8(1+a-2a^2)f(2^2)+(2-10a+17a^2)f(21^2).$$

The remaining roots are single, with the linear components set out below.

TABLE 37
Roots and Principal Components of Frequency

Root.	Principal Linear Component of Frequency.
$(1-a)/2$	$f(21^2)+2f(1^4)$
$(5-2a)/6$	$3f(31)+4f(2^2)+5f(21^2)+6f(1^4)$

The dominant latent root is $(5-2a)/6$, which takes the value $\cdot 8\dot{3}$ when $a = 0$, and only falls to $\cdot 8$ when $a = \cdot 1$. The rate of progress towards homozygosity by selfing tetrasomics is therefore about the same as with sib-matings, or parent-offspring matings of disomics. Polysomy is, therefore, a powerful means of protection against inbreeding. Its occurrence in plants deserves to be considered from this point of view. The protection is strongest where a tends to zero, in the neighbourhood of the centromeres.

The limiting frequency ratio of the digenic genotypes is given by

$$\frac{f(31)}{8+8a-16a^2} = \frac{f(2^2)}{6-6a+9a^2} = \frac{f(31)+f(2^2)}{14+2a-7a^2} = \frac{3f(31)+4f(2^2)}{12(4-a^2)};$$

consequently the proportion heterogenic after n generations of selfing tends to

$$\frac{14+2a-7a^2}{12(4-a^2)} \cdot \left(\frac{5-2a}{6}\right)^n \left\{3f_0(31)+4f_0(2^2)+5f_0(21^2)+6f_0(1^4)\right\}.$$

It will be noticed that the coefficients of the frequencies are the sums of the products of the parts of the corresponding partitions taken two at a time.

23. Selfing Hexasomic Organisms

To follow out the results of selfing in hexasomic organisms, the work is similar, though heavier algebra has to be handled. The generation matrix is presented in sections of three or two columns in the following table (38).

TABLE 38

	$f_0(51)/36$	$f_0(42)/225$	$f_0(3^2)/200$
$f_1(51)$	$18-6\beta-4\beta^2$	$54+30\beta-28\beta^2$	$18+48\beta-18\beta^2$
$f_1(42)$	$9-6\beta+7\beta^2$	$108-42\beta+22\beta^2$	$99-6\beta-9\beta^2$
$f_1(3^2)$	$6\beta-4\beta^2$	$54-6\beta-4\beta^2$	$82-48\beta+18\beta^2$

	$f_0(41^2)/150$	$f_0(321)/3600$	$f_0(2^3)/150$
$f_1(51)$	$36+16\beta-20\beta^2$	$162+528\beta-130\beta^2$	$12\beta+4\beta^2$
$f_1(42)$	$27-14\beta+30\beta^2$	$549+150\beta+253\beta^2$	$9+18\beta+5\beta^2$
$f_1(3^2)$	$24\beta-13\beta^2$	$324+156\beta+8\beta^2$	$9+6\beta+5\beta^2$
$f_1(41^2)$	$39-30\beta-\beta^2$	$486+180\beta-378\beta^2$	$9+30\beta-23\beta^2$
$f_1(321)$	$36-2\beta-10\beta^2$	$1692-672\beta-156\beta^2$	$90-24\beta-18\beta^2$
$f_1(2^3)$	$6-6\beta+8\beta^2$	$378-396\beta+306\beta^2$	$33-42\beta+25\beta^2$

	$f_0(31^3)/1200$	$f_0(2^21^2)/900$
$f_1(51)$	$54+144\beta-54\beta^2$	$48\beta+16\beta^2$
$f_1(42)$	$81-18\beta+161\beta^2$	$27+42\beta+61\beta^2$
$f_1(3^2)$	$108\beta-28\beta^2$	$18+36\beta+20\beta^2$
$f_1(41^2)$	$216-88\beta^2$	$45+150\beta-83\beta^2$
$f_1(321)$	$324+192\beta-164\beta^2$	$288+156\beta-140\beta^2$
$f_1(2^3)$	$84-96\beta+140\beta^2$	$99-114\beta+127\beta^2$
$f_1(31^3)$	$168-96\beta-72\beta^2$	$108+24\beta-132\beta^2$
$f_1(2^21^2)$	$270-252\beta+78\beta^2$	$315-342\beta+123\beta^2$

	$f_0(21^4)/900$	$f_0(1^6)/60$
$f_1(51)$	$24\beta+8\beta^2$	$\cdot$
$f_1(42)$	$9+6\beta+61\beta^2$	$2\beta^2$
$f_1(3^2)$	$24\beta+20\beta^2$	$2\beta^2$
$f_1(41^2)$	$27+90\beta+3\beta^2$	$8\beta^2$
$f_1(321)$	$108+324\beta-144\beta^2$	$24\beta-8\beta^2$
$f_1(2^3)$	$63-90\beta+147\beta^2$	$3-6\beta+11\beta^2$
$f_1(31^3)$	$108+204\beta-312\beta^2$	$36\beta-36\beta^2$
$f_1(2^21^2)$	$405-342\beta+153\beta^2$	$27-18\beta+15\beta^2$
$f_1(21^4)$	$180-240\beta+60\beta^2$	$27-30\beta+3\beta^2$
$f_1(1^6)$	$\cdot$	$3-6\beta+3\beta^2$

TABLE 39

$(1-\beta)^2 \div 20$

$f(1^6)$	$2(9-24\beta+23\beta^2)$	$24(\;3-\;4\beta+\;\;\beta^2)$	$-234+1170\beta-2000\beta^2+1182\beta^3-\;390\beta^4$
$f(21^4)$	$-2(1-\beta)(9+11\beta^2)$	$-4(49-57\beta+28\beta^2)$	$\;\;152-\;598\beta+1170\beta^2-1294\beta^3+\;330\beta^4$
$f(2^21^2)$	$45-42\beta+13\beta^2$	$15(19-10\beta-\;9\beta^2)$	$-138+\;393\beta-\;248\beta^2-\;295\beta^3+\;416\beta^4$
$f(31^3)$	·	$4(279-42\beta-77\beta^2)$	$\;\;738+\;732\beta+\;604\beta^2-2316\beta^3+1042\beta^4$
$f(2^3)$	·	$-3(111-78\beta+47\beta^2)$	$-522-\;\;18\beta+1344\beta^2-2306\beta^3+\;862\beta^4$
$f(321)$	·	$4(\;42+\;9\beta-31\beta^2)$	$1752-4752\beta+5726\beta^2-1824\beta^3+\;458\beta^4$
$f(41^2)$	·	·	$3(2403-1098\beta-\;636\beta^2-\;646\beta^3+297\beta^4)$
$f(3^2)$	·	·	$-8(\;729-\;459\beta-\;273\beta^2+\;277\beta^3+\;74\beta^4)$
$f(42)$	·	·	$10(\;324-\;639\beta+\;588\beta^2-\;107\beta^3+\;10\beta^4)$
$f(51)$	·	·	·

(first value column: $f(1^6)=1$, remainder ·)

$(3-\beta)(1-\beta) \div 15$ $(33-27\beta+2\beta^2)\div 75$

	$(3-\beta)(1-\beta)\div 15$		$(33-27\beta+2\beta^2)\div 75$	
$f(1^6)$	3	$-2(18-39\beta+41\beta^2)$	15	$30(\;87+\;7\beta+\;33\beta^2)$
$f(21^4)$	1	0	9	$2(801+36\beta+334\beta^2)$
$f(2^21^2)$	·	$-2(\;3+16\beta-19\beta^2)$	4	$2(339-41\beta+201\beta^2)$
$f(31^3)$	·	$5(\;9-\;6\beta+\;5\beta^2)$	3	$3(237+\;2\beta+108\beta^2)$
$f(2^3)$	·	$-2(27+\;9\beta-16\beta^2)$	·	$-6(\;27+42\beta-\;32\beta^2)$
$f(321)$	·	$-3-16\beta+19\beta^2$	·	$-5(\;\;9+36\beta-\;34\beta^2)$
$f(41^2)$	·	$2(24-\;7\beta+\;3\beta^2)$	·	$-6(66+11\beta+\;19\beta^2)$
$f(3^2)$	·	·	·	$-18(24+19\beta-\;\;9\beta^2)$
$f(42)$	·	·	·	$-4(27+42\beta-\;32\beta^2)$
$f(51)$	·	·	·	$50(\;9+\;3\beta+\;\;\;\beta^2)$

$(7-3\beta) \div 10$ $(27-4\beta) \div 30$

	$(7-3\beta)\div 10$	$(27-4\beta)\div 30$
$f(1^6)$	10	15
$f(21^4)$	8	14
$f(2^21^2)$	6	13
$f(31^3)$	5	12
$f(2^3)$	4	12
$f(321)$	3	11
$f(41^2)$	2	9
$f(3^2)$	·	9
$f(42)$	·	8
$f(51)$	·	5

The ten latent roots are all functions of β, positive, rational and diminishing as β increases. The smallest root is quadruple for all values of β, the two next larger are double roots, and only the two largest are single roots. Table 39 shows the five values of the latent roots, and the corresponding principal components of frequency. For each value one component of frequency is a very simple form independent of β, the coefficients being the sums of the products of the parts of the corresponding partitions taken 2, 3, 4, 5 or 6 at a time.

The determination of the other principal components offers the difficulty that they are algebraically indeterminate. *E.g.* the third component given for the root $(1 - \beta^2)/20$ might be altered by an arbitrary multiple of the second component. To render it determinate by assigning an arbitrary value, zero, to one of the pentasomic frequencies, would yield a non-integral solution with a denominator quadratic in β. To obtain integral solutions was troublesome, and I do not suppose the form in which I have presented these coefficients is the most intelligible possible.

24. Sib-matings with Tetrasomic Inheritance

Among the many problems of inbreeding which can only be treated cursorily are those in which two polyploids are mated. The number of mating types comes in these cases to be very large. Although, as we have seen, there are only five distinct partitions of the number 4, and so only five types in tetrasomic inheritance corresponding with homozygotes and heterozygotes in disomic inheritance, yet the number of mating types is found to be sixty-eight, and, as a combinatorial problem, these are difficult to enumerate. Classified according to the number of genes present, there are found (Table 40, overleaf).

TABLE 40

Number of Genes.	Number of Mating Types.	Number of Genes.	Number of Mating Types.
1	1	5	13
2	8	6	6
3	17	7	2
4	20	8	1

Although, therefore, the complete system is hierarchical, the diagonal blocks in a complete enumeration would have up to twenty rows and columns. We may therefore confine attention to cases in which only two genes are present, and we have only to deal with an 8×8 matrix.

Even in this case it would be difficult to handle the matrix in general when double reduction is taken into account. The eight matings involving two genes may be represented as follows :—

TABLE 41

Genetical Formula.	Brief Designation.	Genetical Formula.	Brief Designation.
$a_4 \times Aa_3$	$N \times S$	$a_4 \times A_3 a$	$N \times S'$
$Aa_3 \times Aa_3$	$S \times S$	$A_2 a_2 \times A_2 a_2$	$D \times D'$
$a_4 \times A_2 a_2$	$N \times D$	$Aa_3 \times A_3 a$	$S \times S'$
$Aa_3 \times A_2 a_2$	$S \times D$	$a_4 \times A_4$	$N \times N'$

If we enquire for any mating types to what genotypes the offspring will belong, we may use the gametic frequencies of Table 30 and find :—

TABLE 42

	$a_4 \times Aa_3$	$Aa_3 \times Aa_3$	$a_4 \times A_2 a_2$	$Aa_3 \times A_2 a_2$	$a_4 \times A_3 a$	$A_2 a_2 \times A_2 a_2$	$Aa_3 \times A_3 a$	$a_4 \times A_4$
a_4	$2+ a$	$4+4a+ a^2$	$1+2a$	$2+ 5a+ 2a^2$	a	$1+ 4a+ 4a^2$	$2a+ a^2$	·
Aa_3	$2-2a$	$8-4a-4a^2$	$4-4a$	$10- 2a- 8a^2$	$2-2a$	$8+ 8a-16a^2$	$4 -4a^2$	·
$A_2 a_2$	a	$4-4a+6a^2$	$1+2a$	$10-10a+12a^2$	$2+ a$	$18-24a+24a^2$	$8-4a+6a^2$	1
$A_3 a$	·	$4a-4a^2$	·	$2+ 6a- 8a^2$	·	$8+ 8a-16a^2$	$4 -4a^2$	·
A_4	·	a^2	·	$a+ 2a^2$	·	$1+ 4a+ 4a^2$	$2a+ a^2$	·
	4	16	6	24	4	36	16	1

Next, if pairs of sibs are chosen at random for mating, the frequencies with which different mating types will be made up will in three cases be quadratic in a, as in selfing tetraploids, but in four more cases will be of the fourth degree. For example, from $N \times S$ and $S \times S$ for comparison, we shall have :—

TABLE 43

	$\dfrac{N \times S}{16}$	$\dfrac{S \times S}{256}$
$N \times S$	$8-4a-4a^2$	$64+32a-48a^2-32a^3-16a^4$
$S \times S$	$4-8a+4a^2$	$64-64a-32a^2 \qquad +32a^4$
$N \times D$	$4a+2a^2$	$32 \quad +32a^2+32a^3+24a^4$
$S \times D$	$4a-4a^2$	$64-64a+32a^2+64a^3-96a^4$
$N \times S'$	.	$32a+16a^2-32a^3-16a^4$
$D \times D$	a^2	$16-32a+64a^2-48a^3+36a^4$
$S \times S'$	.	$64a-96a^2 \qquad +32a^4$
$N \times N'$	.	$8a^2+ 8a^3+ 2a^4$

The determinant will therefore be a very troublesome one, being not only of the eighth degree in λ but of the twenty-second degree in a. We have learnt, however, from the case of selfing with tetrasomic inheritance that the effect of double reduction in increasing the speed with which the homogenic condition is established, though perceptible, was not there large. Consequently, it will be useful to consider in more difficult problems the case $a = 0$, appropriate to loci not far from the centromere.

The 8×8 λ-matrix is then :—

TABLE 44

	$\dfrac{N \times S}{4}$	$\dfrac{S \times S}{16}$	$\dfrac{N \times D}{36}$	$\dfrac{S \times D}{144}$	$\dfrac{N \times S'}{4}$	$\dfrac{D \times D}{648}$	$\dfrac{S \times S'}{8}$	$\dfrac{N \times N'}{1}$
$N \times S$	$2-2\lambda$	4	8	10	.	16	.	.
$S \times S$	1	$4-16\lambda$	16	26	1	64	1	.
$N \times D$	.	2	$2-36\lambda$	10	.	36	.	.
$S \times D$	.	4	8	$60-144\lambda$	2	288	4	.
$N \times S'$	.	.	.	2	-4λ	16	.	.
$D \times D$	.	1	1	25	1	$162-648\lambda$	2	1
$S \times S'$	.	.	.	10	.	64	$1-8\lambda$	.
$N \times N'$	.	.	.	.	.	1	.	$-\lambda$

Direct evaluation and simplification of the determinant shows it to have a constant factor, -256, and the factor in λ of the eighth degree

$$1+19\lambda-2434\lambda^2-3148\lambda^3+703680\lambda^4+840960\lambda^5-35,168256\lambda^6$$
$$+85,847040\lambda^7-53,747712\lambda^8.$$

When such an expression has factors it is not easy to recognise them. Since, however, the eight numerical values of λ which reduce it to zero are all real, they may be calculated with sufficient accuracy, and their reciprocals examined, whereupon it appears that there is one linear factor

$$1-36\lambda,$$

two quadratic factors

$$1+10\lambda-12\lambda^2$$

and

$$1+52\lambda-288\lambda^2,$$

and one cubic factor

$$1-7\lambda-240\lambda^2+432\lambda^3.$$

The latent roots of the λ-matrix in order of absolute magnitude are therefore :—

TABLE 45

	λ	$1/\lambda$
$(5+\sqrt{37})/12$	·92356,35442	1·08276,2530
	·57669,29002	1·73402,5163
$(13+\sqrt{241})/144$	·19808,45465	5·04834,9393
$(5-\sqrt{37})/12$	$-$·09023,02108	-11·08276,2530
	$-$·07479,98483	-13·36901,1070
	·05366,25037	18·63498,5907
$1/36$	·02777,77778	36·
$(13-\sqrt{241})/144$	$-$·01752,89910	-57·04834,9392

From the reciprocals it is easy to see which roots belong to quadratic factors.

For the dominant root

$$-\log\lambda = \cdot079515,6737$$
$$-1/\log\lambda = 12\cdot56714.$$

Compared with sib-mating in diploids (for which $\log_e 1/\epsilon = \cdot 21193,53554$), the effect of tetraploidy is to retard the approach to the homogenic condition, in the ratio $2\cdot 66533$, so that the time needed to produce equivalent effects is increased in this ratio. For selfing the corresponding ratio is $3\cdot 80178$, both these being maximal values accurate when a is zero. It is evident that in a self-sterile species in which the closest inbreeding commonly practised is sib-mating, tetrasomic inheritance affords a powerful additional protection against the immediate injury due to homozygosity.

In the case of irrational roots encountered in previous examples it has been not inconvenient to express the complexities and the corresponding frequencies exactly as linear functions of the dominant root. When this root is derived from a quadratic equation this is always possible, but, as the present example shows, it may have little real advantage over a direct numerical solution.

The complexities do indeed stand in ratios to each other expressible as simple functions of λ, but the frequencies do not, so that when the complexities are expressed in terms of the average surviving mating type, a very cumbrous factor is involved. The absolute complexity of the mating type $N \times S$ in fact comes to

$$486,438476\lambda + 41,808623 \div 866,434032$$

and the others will have similarly complex expressions. The simple ratios and the numerical values for the absolute complexities are shown in Table 46.

The frequencies to which after long sib-mating these mating types will tend do not stand in simple ratios one to another. The exact expressions are shown in Table 47.

Using these expressions it is easily found that a parent-offspring mating intercalated in a sib-mating series reduces the complexity in the ratio $\cdot 937736$, and is equivalent in sib-matings to only $\cdot 8085$ of a generation.

TABLE 46

Complexities of Eight Mating Types

Mating Type.	In Terms of the Least Heterogenic Mating.	Absolute Numerical Value.
$N \times S$	1	·56677
$S \times S$	$4\lambda - 2$	·96025
$N \times D$	$4(5 - 4\lambda)/3$	·98674
$S \times D$	$(8\lambda - 1)/3$	1·20693
$N \times S'$	$17 - 16\lambda$	1·25991
$D \times D$	$8(2\lambda - 1)/3$	1·28033
$S \times S'$	$6 - 4\lambda$	1·30682
$N \times N'$	$32 - 32\lambda$	1·38629

TABLE 47

Frequencies of Eight Mating Types

Mating Type.	Exact Relative Frequency.	Numerical Value, per cent.
$N \times S$	$135909,294720\lambda - 124073,000272$	23·02019
$S \times S$	$-945299,974464\lambda + 874534,242688$	23·68444
$N \times D$	$579153,470472\lambda - 534471,958836$	6·56759
$S \times D$	$354573,891072\lambda - 325593,325440$	29·86207
$N \times S'$	$-100934,657856\lambda + 93267,985744$	·76977
$D \times D$	$171360,906264\lambda - 157508,224704$	11·99545
$S \times S'$	$-191846,111232\lambda + 177438,716288$	4·08044
$N \times N'$	$- 2916,818976\lambda + 2695,128323$	·02004
Total	$\cdot$ 6289,563791	99·99999

25. Parent-offspring Matings with Tetrasomic Inheritance

With parent-offspring matings distinction must be made when the genotypes are not symmetrically related between that of the parent and that of the offspring. If, however, double reduction is ignored, the mates, if a parent and offspring, must have at least two genes in common; for this reason only nine of the twelve possible mating types occur in a regular series of parent-offspring matings, for $N \times S'$, $S' \times N$ and $N \times N'$ are excluded, and will only occur in an irregular mating.

The λ-matrix for the nine possible mating types is :-

TABLE 48

λ-matrix for Parent-offspring Inbreeding

O P	$\dfrac{N\times S}{2}$	$\dfrac{S\times N}{2}$	$\dfrac{N\times D}{6}$	$\dfrac{S\times S}{4}$	$\dfrac{S\times D}{12}$	$\dfrac{D\times N}{6}$	$\dfrac{D\times S}{12}$	$\dfrac{S\times S'}{4}$	$\dfrac{D\times D}{18}$
$N\times S$	-2λ	1	.	1	1	.	.	.	.
$S\times N$	1	-2λ	4	.	.	.	.	.	.
$N\times D$	.	.	-6λ	.	.	1	1	.	1
$S\times S$	.	1	.	$2-4\lambda$	5	.	.	1	.
$S\times D$	.	.	.	.	-12λ	4	6	.	8
$D\times N$	.	.	1	.	.	-6λ	.	.	.
$D\times S$	.	.	.	1	5	.	-12λ	2	.
$S\times S'$	.	.	.	.	1	.	.	$1-4\lambda$	.
$D\times D$	.	.	.	.	.	1	5	.	$9-18\lambda$

Expressed as a polynomial in λ the determinant comes to

$$576\lambda^2(1-26\lambda+116\lambda^2+548\lambda^3-3432\lambda^4-144\lambda^5+12960\lambda^6-10368\lambda^7).$$

There is a double root at zero, and the remaining seven are the roots of the three factors

$$(1-6\lambda)(1-16\lambda-60\lambda^2+144\lambda^3)(1-4\lambda-8\lambda^2+12\lambda^3).$$

The roots are all real, and in order of absolute magnitude are

·92893,00976
·58604,24264
−·45809,41990
−·22260,73015
·19583,07681
·16666,66667
·05323,15418

and twice

0·0.

For the dominant root,

$$\log 1/\lambda = \quad ·07372,17879$$
$$1/\log 1/\lambda = 13·56451.$$

The speed of approach to the homogenic condition is thus slower than for sib-mating by 7·637 per cent., and the time taken to make equal progress is 2·87480 times as great as for disomic organisms bred in the same way.

The numerical values of the frequencies and complexities corresponding with the dominant root are :—

<div align="center">TABLE 49</div>

	Frequency, per cent.	Complexity.
$N \times S$	13·4305	·396705
$S \times N$	8·9159	·737023
$N \times D$	2·3505	·743826
$S \times S$	26·8215	·972579
$S \times D$	15·7518	1·158530
$D \times N$	·4217	1·197566
$D \times S$	15·3245	1·271972
$S \times S'$	1·9334	1·294878
$D \times D$	15·0502	1·296770
Total	100·0000	

The corresponding complexities to be ascribed to the three missing mating types, if one of these were introduced as an irregular mating, are easily obtained. For example, the mating $N \times S'$ leads with equal frequency to $D \times N$ and $S \times N$; its complexity is therefore

$$\frac{1}{2\lambda}(1·197566 + ·737023) = 1·041300 ;$$

for $S' \times N$ we find equally

$$\frac{1}{2\lambda}(1·271972 + 1·294878) = 1·381616$$

and for $N \times N'$

$$\frac{1}{\lambda}(1·197566) \qquad = 1·289181.$$

Evidently $S' \times N$ is the most complex mating type from which to start a series of parent and offspring

matings ; such organisms could not indeed, without double reduction, be related as parent and off-spring, but any such mating in which the simplex member was to be used a second time would, as a progenitor of a parent-offspring line, receive this designation.

There is thus no special difficulty in evaluating the effect of a sib-mating intercalated in a continuous series of parent-offspring matings. This is found to diminish the complexity in the ratio ·932671, and has 94·55 per cent. of the effect of a parent-offspring mating. In both series, therefore, an irregularity somewhat delays progress, although in a continuous series, sib-matings in tetrasomic organisms are more efficient than parent-offspring matings.

26. Double Cousin Inbreeding

In a good many bisexual organisms sib-mating is wholly or largely precluded by unisexual families. In man an instinctive reluctance to the mating of near kin has apparently led to such matings being forbidden by social custom. Cousin and double cousin marriages are, however, widely allowed, and among many peoples, both civilised and uncivilised, are so customary as to be almost obligatory. A widespread convention is for a son to marry his mother's brother's daughter, and for a woman to marry her father's sister's son. Such a rule allows of the perpetuation of a line with only two matings in each generation and, although it is probable that irregularities must often occur, owing to the failure of offspring of one sex or the other, it is by no means certain that some peoples have not been inbred approxi-mately as closely as such a rule would imply for very long periods ; for the custom is widely diffused and therefore presumably very ancient. Tests of the

homozygosity of common genes, such as those determining the blood groups, M and N, should supply a useful measure of the extent to which inbreeding has actually progressed in such cases.

As was shown in a previous example, when only two genes are considered, the number of types of pairs of matings is thirteen. This gives a 12×12 matrix which is just not unmanageable. Indeed the practical difficulty of this example is more due to a second circumstance, which has not been encountered in previous cases, namely that many of the latent roots involve an imaginary component, and are mathematically speaking complex numbers. The numerical extraction of complex roots of an equation is much more troublesome than that of the real though irrational roots, with which we have principally been concerned.

Using the designations previously adopted for these pairs of matings, and following out the genetic consequences of each, we find the following 12×12 generation matrix, ignoring the homogenic mating type tt :—

<div align="center">TABLE 50</div>

	$\dfrac{tu}{4}$	$\dfrac{uu}{16}$	$\dfrac{tv}{16}$	$\dfrac{tw}{1}$	$\dfrac{uv}{64}$	$\dfrac{vv}{128}$	$\dfrac{uw}{4}$	$\dfrac{vw}{8}$	$\dfrac{tu'}{4}$	$\dfrac{uu'}{16}$	$\dfrac{ww}{1}$	$\dfrac{tt'}{1}$
tu	2	4	4	·	6	8	·	·	·	·	·	·
uu	1	4	4	1	10	16	1	1	1	2	·	·
tv	·	2	·	·	4	8	·	·	·	·	·	·
tw	·	·	2	·	2	4	·	·	·	·	·	·
uv	·	4	·	·	16	32	2	4	·	4	·	·
vv	·	1	·	·	4	8	1	2	·	1	1	·
uw	·	·	4	·	8	16	·	·	2	4	·	·
vw	·	·	·	·	4	8	·	·	·	2	·	·
tu'	·	·	·	·	2	8	·	·	·	·	·	·
uu'	·	·	·	·	6	16	·	1	·	2	·	·
ww	·	·	1	·	1	2	·	·	1	1	·	1
tt'	·	·	·	·	·	1	·	·	·	·	·	·

Apart from the numerical factor -128, the deter-

minant may be reduced to the following expression of the twelfth degree :—

$$
\begin{aligned}
&1\\
&+2\lambda\\
&-156\lambda^2\\
&-2184\lambda^3\\
&-6912\lambda^4\\
&+117760\lambda^5\\
&+147968\lambda^6\\
&+1,224704\lambda^7\\
&-11,190272\lambda^8\\
&-1,703936\lambda^9\\
&-19,922944\lambda^{10}\\
&+159,383552\lambda^{11}\\
&-134,217728\lambda^{12}.
\end{aligned}
$$

After evaluating the real roots and two at least of the imaginary pairs, it appears that the polynomial in λ is the product of four rational factors, one of the first degree, two cubics and one quintic, as shown below.

TABLE 51

Linear Factor.	Cubic Factors.		Quintic Factor.
1	1	1	1
-16λ	$+2\lambda$	$+8\lambda$	$+8\lambda$
	$+4\lambda^2$	$+64\lambda^2$	$-32\lambda^2$
	$-8\lambda^3$	$-512\lambda^3$	$-64\lambda^3$
			$-1024\lambda^4$
			$+2048\lambda^5$

The cubic factors both have one real and one pair of complex roots. The two cubics present the peculiarity that substituting λ for 4λ reduces the second to the first ; consequently each root of the second equation is one quarter of the corresponding root of the first equation. For the first cubic the roots are

Real Root.	Complex Roots.	
$\cdot91964,33776$	$-\cdot20982,16888$	or $\cdot36867,63529$
	$\pm\cdot30314,53646\,i$	exp $i\ 124°\cdot6889974.$

The real root here is the dominant root of the entire matrix ; for this value

$$
\begin{aligned}
\log_e \lambda &= \quad\cdot08376,93187,5\\
1/\log_e \lambda &= 11\cdot93754,4854.
\end{aligned}
$$

Inbreeding by continued double cousin mating in diploids is thus more rapid than sib-mating in tetraploids by 5·140 per cent., and 12·777 per cent. more rapid than parent-offspring mating of tetraploids, in the absence of double reduction, while it is 22·942 per cent. slower than the self-fertilisation of hexaploids with the same restriction. The complex roots have been specified above in the two forms

$$x + iy = r \exp i\theta$$

where θ has been given in degrees. The amplitude r, which is positive, is the rate of decay of the absolute value of the corresponding complex linear component of the frequencies, and the angle is the increment in each generation of its argument. Thus if

$$L + iM$$

is the linear component of the frequencies corresponding with a complex root, L and M being real, we have the relation

$$L_n + iM_n = r^n e^{in\theta}(L_0 + iM_0) ;$$

equating real and imaginary parts, we find that

$$L_n = r^n(L_0 \cos n\theta - M_0 \sin n\theta)$$
$$M_n = r^n(L_0 \sin n\theta + M_0 \cos n\theta)$$

are a conjugate pair of real equations, expressing the frequencies after n generations of inbreeding in terms of the initial values of these conjugate components.

By reason of the relationship between the roots of the second cubic and those of the first, the three roots of this factor are easily derived from those of the first, and the angle of the complex pair of roots is unchanged.

The quintic factor has three real roots and one complex pair, namely :—

TABLE 52

Real Roots.	Complex Roots.		
·57642,13329	−·09044,90722,7	or	·22133,95947
·19372,87567	±·20201,53001,5 i		exp i 114°·1196743
−·08925,19450,5			

Of the six real roots, therefore, five are positive and one negative, while the real parts of all six complex roots are negative. The absolute values in order of magnitude are :—

TABLE 53

Source					
First cubic	.	.	.	.	·91964
Quintic	.	.	.	.	·57642
First cubic, pair	.	.	.	.	·36868
Second cubic	.	.	.	.	·22991
Quintic, pair	.	.	.	.	·22134
Quintic	.	.	.	.	·19373
Second cubic, pair	.	.	.	.	·09216,9
Quintic	.	.	.	.	·08925,2
Linear factor	.	.	.	.	·06250,0

The numerical values of the frequencies and complexities corresponding with the dominant root are :—

TABLE 54

Matings.	Symbol.	Frequency, per cent.	Complexity.
$aa \times aa$, $aa \times ab$	tu	23·28074	·5719006
$aa \times ab$, $aa \times ab$	uu	23·97967	·9599771
$aa \times aa$, $ab \times ab$	tv	5·23339	·9895201
$aa \times aa$, $aa \times bb$	tw	1·69835	1·0438580
$aa \times ab$, $ab \times ab$	uv	20·83017	1·1937726
$ab \times ab$, $ab \times ab$	vv	8·21620	1·2437442
$aa \times ab$, $aa \times bb$	uw	7·18706	1·2481106
$ab \times ab$, $aa \times bb$	vw	2·53799	1·2980821
$aa \times aa$, $ab \times bb$	tu'	1·26620	1·3071965
$aa \times ab$, $ab \times bb$	uu'	4·14916	1·3276251
$aa \times bb$, $aa \times bb$	ww	1·55126	1·3524201
$aa \times aa$, $bb \times bb$	tt'	·06980	1·4705919

99·99999

REFERENCE

(6) R. A. FISHER and K. MATHER (1943). The inheritance of style length in *Lythrum salicaria*. Ann. Eugen., xii. 1, 1-23.

APPENDIX A

SPECIES BEARING ONE OFFSPRING AT A BIRTH

27. Model Mating Systems

In many species, and those among the slowest breeders, one offspring only is to be expected at a birth. The use of successive sib-matings is then particularly time-consuming, for, whichever may be the sex of the first-born, a second mating will certainly be required, and this will produce an offspring of the right sex in only half the trials. The frequency of a delay of 1, 2, 3, ... birth intervals after the first birth is given by the geometric series

$$\begin{array}{cccccc} 1 & 2 & 3 & \cdots & \cdots & \cdots \\ \tfrac{1}{2} & \tfrac{1}{4} & \tfrac{1}{8} & \cdots & \cdots & \cdots \end{array}$$

from which it is easily seen that the average delay, beyond the time needed for the first birth, will be two birth intervals.

The alternative of using parent-offspring matings, using each animal twice only, which, as has been seen, is equally efficacious in securing homozygosis, is manifestly preferable for such species, since young of the right sex will be supplied by the first birth in half the cases, by the second birth in quarter, and so on. Such a system therefore saves time, as compared with sib-mating, at the rate of exactly one birth interval in each generation.

In cases, however, in which the first, or the first two births produce offspring of the wrong sex, it may well be asked whether use cannot be made of these to advance the progress of the line more expeditiously than by waiting passively for the appearance of an

offspring of the desired sex. Several alternative possibilities present themselves, some of which are discussed below as examples of the manner in which such operational policies may be investigated.

MODEL C.—If the first offspring is always used with the parent of opposite sex, the generation matrix is

				$aa \times ab$	$ab \times ab$
				u_0	v_0
u_1	.	.	.	$\frac{1}{2}$	$\frac{1}{2}$
v_1	.	.	.	$\frac{1}{4}$	$\frac{1}{2}$.

The equation for λ is

$$8\lambda^2 - 8\lambda + 1 = 0.$$

The dominant root

$$\lambda = \cdot 85355,33906 = \tfrac{1}{4}(2 + \sqrt{2})$$
$$-\log_e \lambda = \cdot 15835,88971$$
$$-1/\log_e \lambda = 6\cdot 31476992.$$

If g is the ratio of the minimal generation time to the time interval between births, the time taken by this method to attain a standard degree of inbreeding may be compared with that required by other methods :—

(C) First offspring method . . . $(6\cdot31477)g$
(B) Alternate parent-offspring method . $(4\cdot71842)(g+1)$
(A) Sib-matings . . . . $(4\cdot71842)(g+2)$.

It is obvious that some parent-offspring method is quicker than the sib-method with animals bearing only one at a birth ; also, if g is less than $2\cdot9557$, the use of the first offspring is quicker than the policy of waiting for one of the sex wanted.

The comparatively high value of the root, $\cdot85355$, shows that the use of the same animal in more than two generations is somewhat disadvantageous. Reiterated use is also avoidable, for a possible policy would be to use the same parent no more than three times,

and to continue every third mating until an offspring of suitable sex appears. Although, as will appear, this method can be improved on, it is in many circumstances better than those discussed above. The method of calculating its speed is also of some interest.

In this case (model D) we must distinguish six possible types of mating, in three of which the parent is used for the second time, and in three for the third time. Writing the genotype of the parent first, these are :—

$$ab \times aa \qquad\qquad aa \times ab \qquad\qquad ab \times ab$$
$$u \qquad\qquad\qquad\quad v \qquad\qquad\qquad\quad w$$

using the parent for the second time, and u', v', w' when used for the third time. The generation matrix is then found to be :—

TABLE 55

	$u_0/4$	$v_0/4$	$w_0/4$	$u'_0/2$	$v'_0/2$	$w'_0/2$
u_1	-4λ	1	1	.	1	1
v_1	1	-4λ	.	1	.	.
w_1	.	1	$1-4\lambda$	.	1	1
u'_1	1	.	1	-2λ	.	.
v'_1	.	1	.	.	-2λ	.
w'_1	1	.	1	.	.	-2λ

Two of the roots for λ are zero, and the remainder are roots of the quartic equation

$$32\lambda^4 - 8\lambda^3 - 10\lambda^2 - 3\lambda - 1 = 0.$$

The dominant root is

$$\lambda = \cdot 82358,93624$$
$$-\log_e \lambda = \cdot 19408,32199$$
$$-1/\log_e \lambda = 5\cdot 14253.$$

Since the third use of a parent only occurs when on the second use the offspring is of the wrong sex, only one third of the matings will be of this kind, after which an offspring of the right sex must be awaited.

The time needed for a standard amount of inbreeding is therefore

$$(5 \cdot 14253)(g + \tfrac{1}{3})$$

and this is less than the value for alternate parent-offspring matings, if g is less than $6 \cdot 91$, a value exceeded by man, but I think not by other animals. It is also less than the value for the method using always the first offspring when g exceeds $1 \cdot 48$, which also must generally be the case.

This method is not, however, to be recommended in practice since, for example, when the parent is a male, he might have a second offspring by his daughter, before the first by his granddaughter. If this turns out to be a male, either a mating to his mother or to his sister will become possible. If the granddaughter bears a son, he with his mother will carry on the line ; but if a daughter, she could be mated to any brother of her mother.

When a male is mated repeatedly with a daughter and a granddaughter with a view to using a male born to one of the matings with his mother, the order of preference among the males born will influence both the retardation of the matings ·relative to the minimal generation time, and the proportion of cases in which one animal is used three times in succession. Only an exact calculation can show the simultaneous effect of both influences.

Let us suppose that when a male is bred to a daughter and granddaughter the order of preference is as follows :—

Offspring of Daughter.	Offspring of Granddaughter.
1	
2	
4	3 ·
6	5
.	7
.	.
.	.
.	.

while a female is only mated to a grandson if the first two born to her son are both male. Then the fraction of males used in a third generation is

$$\frac{1}{8}\left(1 + \frac{1}{4} + \frac{1}{4^2} + \cdots\right) = \frac{1}{6},$$

while that of females is

$$\frac{1}{8}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right) = \frac{1}{4}.$$

In all, therefore,

$$\tfrac{1}{2}(1/6 + 1/4) = \tfrac{5}{24}$$

is the frequency of carrying on the line by third matings. Further, for males the average delay is

$$(\tfrac{1}{4} + \tfrac{1}{32})\{1 + 2(\tfrac{1}{4}) + 3(\tfrac{1}{4})^2 + \cdots\} = \tfrac{1}{2}$$

and for females

$$\tfrac{1}{4} + \tfrac{1}{16}\{1 + 2(\tfrac{1}{2}) + 3(\tfrac{1}{2})^2 + \cdots\} = \tfrac{1}{2}.$$

The situation is then that of twelve second matings by males, two are carried on by a third generation using the first born, and ten as second matings of the mother, with a total delay of $5\frac{1}{8}$ birth intervals; the two third generation matings being continued as second matings to the female with a total delay of $\frac{2}{3}$ of a birth interval. While of twelve second matings of females, three are carried on by third generation matings to the first born, and nine as second generation matings of the father, with a total delay of three birth intervals; the three third generation matings being continued as second matings to the male, with a total delay of three birth intervals. Thus the total delay is twelve birth intervals in twenty-nine generations, or the average time of a generation is $g + \frac{12}{29}$.

The generation matrix depends only on the propor-

tion of third matings used, irrespective of the time elapsed ; multiplied by 48 it is :—

TABLE 56

-48λ	19	19	.	24	24
19	-48λ	.	24	.	.
.	19	$19-48\lambda$	.	24	24
5	.	5	-48λ	.	.
.	5	.	.	-48λ	.
5	.	5	.	.	-48λ

From this the equation for λ is found to be

$$4608\lambda^4 - 1824\lambda^3 - 1202\lambda^2 - 285\lambda - 25 = 0$$

of which the dominant root is

$$\lambda = \cdot 81743,52344$$
$$- \log_e \lambda = \cdot 20158,36035$$
$$-1/\log_e \lambda = 4\cdot 960720925.$$

The time taken on this system (model E) to make one unit of progress is

$$4\cdot 96072(g+\tfrac{12}{29}) ;$$

this is generally, but only slightly, shorter than the time needed in the first model case involving third matings. By using, however, a different schedule of preference (model F)

	I			I	
	3			3	
Males	5	2	Females		2
	.	4			4
	.	6			5
	.				.
	.				.

the delay is reduced to $\tfrac{17}{65}$ birth intervals, at the expense of using an increased proportion of third matings. This, of course, increases λ, which is now the largest root of

$$18432\lambda^4 - 5952\lambda^3 - 5186\lambda^2 - 1581\lambda - 289 = 0$$
$$\lambda = \cdot 82111,28303$$
$$- \log_e \lambda = \cdot 19709,47486$$
$$-1/\log_e \lambda = 5\cdot 073701898.$$

H

Table 57 shows that in the range of g considered the speed of this method is 31 to 80 per cent. higher than that of sib-matings. In all special cases, other methods will suggest themselves ; it is in order that these may be scrupulously examined before adoption that the full examples set out above have been presented.

TABLE 57

Table of Time, in Birth Intervals, required to Make Unit Advance in Inbreeding, Using Different Methods

$g=$	1·6	2·0	2·4	2·8	3·2	3·6	4·0	
Sib-mating (A)	16·986	18·874	20·761	22·648	24·536	26·423	28·311	$4\cdot71842\ (g+2)$
Alternate parents (B)	12·268	14·155	16·043	17·390	19·817	21·705	23·592	$4\cdot71842\ (g+1)$
First born (C)	10·104	12·630	15·155	17·681	20·207	22·733	25·259	$6\cdot31477\ g$
Model D	9·942	11·999	14·056	16·113	18·170	20·227	22·284	$5\cdot14253\ (g+\tfrac{1}{3})$
Model F	9·445	11·474	13·504	15·533	17·563	19·592	21·622	$5\cdot07370(g+\tfrac{17}{6h})$

28. Time Criterion for Choice of Mating

An objection to any schedule of preference based on birth order is that it takes no account of the actual deviations from average times, which though they cannot be foreseen are known to the experimenter who makes the choice. To take proper account of the circumstances we should need a rule of the form, " The offspring of a third mating is to be preferred to an offspring of the preceding second mating if born not more than x days later."

Since third matings will not, to judge by the results of the preceding section, need to be very numerous, we may, as an approximation, treat the third matings as exceptional irregularities in a series of alternate

parent-child matings. For such a series the matrix :—

TABLE 58

	$u_0/2$	$v_0/2$	$w_0/2$
u_1 . . .	-2λ	1	1
v_1 . . .	1	-2λ	.
w_1 . . .	.	1	$1-2\lambda$

has a dominant root

$$\lambda = \epsilon = \tfrac{1}{4}(\sqrt{5}+1),$$

with frequencies

$$\frac{f(u)}{2\epsilon} = f(v) = (2\epsilon-1)f(w) = \frac{f(w)}{2\epsilon}$$

and complexities

$$c(u) = \frac{c(v)}{2\epsilon} = (2\epsilon-1)c(w) = \frac{c(w)}{2\epsilon}.$$

The expected effect of a third mating is given by the matrix :—

TABLE 59

	u_0	v_0	w_0
u_1 . . .	$\tfrac{1}{2}$	.	$\tfrac{1}{2}$
v_1 . . .	.	$\tfrac{1}{2}$	.
w_1 . . .	$\tfrac{1}{2}$	.	$\tfrac{1}{2}$

so that the reduction in complexity due to a third mating is

$$\tfrac{1}{2}\{f(u)+f(w)\}\{c(u)+c(w)\}+\tfrac{1}{2}f(v)c(v)$$
$$\div f(u)c(u)+f(v)c(v)+f(w)c(w),$$

which may be reduced to

$$\frac{4\epsilon+3}{4\epsilon+4} = \frac{1}{10}(7+2\epsilon) = \cdot86180,33989.$$

The negative natural logarithm is $\cdot14872,81098$ or $70\cdot176$ per cent. of that of a generation of continuous alternate-parent mating. This suggests that when third generation matings are sufficiently rare they will be preferred if the offspring of the right sex born from

the third generation mating is not younger than that born to the second generation mating by more than 70 per cent. of the observed average generation length. If, for example, this average were found to be 890 days, we should prefer the offspring of the third mating to that of the second if born not more than 623 days later. In this way allowance is made for the actual ages of the animals available.

THE EFFICACY OF SELF-STERILITY MECHANISMS AMONG HERMAPHRODITES IN DIMINISHING UNIONS BETWEEN NEAR RELATIONS

The three chief methods of avoiding self-fertilisation in hermaphrodite plants are distylism, tristylism and a system of self-sterility allelomorphs.

All of these are effective in preventing self-fertilisation, but unequally effective in lowering the frequency of matings between near relatives. With distyly, all progenies resemble the general population in giving the two phenotypes in equal numbers ; consequently neither sib-matings nor parent-offspring matings are in any degree diminished by this mechanism. The same applies to diœcious organisms.

In two cases the inheritance of the polymorphic variation of a tristylic species has now been elucidated. The genetic analysis reveals the number of genotypes giving each form of flower, and subject to equal viability and fertility of the different legitimate crosses, the frequencies of the different genotypes of an open-pollinated population, the frequencies of the different legitimate matings, and the frequencies of the genotypes produced by each. On the basis of such an analysis we can therefore infer, subject to the simplifying conditions imposed, the effects of the system on the frequencies of matings between individuals of chosen degrees of kinship.

In *Oxalis valdiviensis* (7) there is disomic inheritance involving two Mendelian factors, rather closely linked. There are seven genotypes produced by legitimate pollination of which one is Long-styled, two are Mid-styled and four Short-styled. The frequencies⁻ in an

open-pollinated population in genetic equilibrium are
shown, with their genetic symbols, in the following table.

TABLE 60

Symbol.				Frequency.	Style Type.
$ms\ /ms$	.	.	.	1	L
$Ms\ /ms$	.	.	.	$4\sqrt{3}-6$	M
$Ms\ /Ms$	.	.	.	$7-4\sqrt{3}$	M
$mS\ /ms$	.	.	.	$4-2\sqrt{3}$	S
MS/ms	.	.	.	$3\sqrt{3}-5$	S
$mS\ /Ms$	.	.	.	$3\sqrt{3}-5$	S
MS/Ms	.	.	.	$7-4\sqrt{3}$	S

For each style type the frequencies given above add
to unity; they are therefore referrable to a total
population of three plants.

Legitimate matings among these seven genotypes,
without distinguishing reciprocal matings, which are
equivalent, number fourteen. Of these two are between
Long and Mid, four between Long and Short, and
eight between Mid and Short. The frequencies of the
different matings will be three times the product of the
frequencies of the two genotypes involved, or one third
of the product of the values given above. The frequencies
expressed in reference to a total of three matings, are
as shown in Table 61.

The proportion of parent-offspring matings which
would be illegitimate is half the sum of the frequencies
in the offspring of the parental phenotypes; for
sib-matings it is the sum of the squares of the three
phenotypic frequencies. Parallel matings of double
heterozygotes in coupling and repulsion which are of
equal frequency may conveniently be averaged. It
will be noticed that linkage makes no difference to the
general frequency of parent-offspring matings, but that
it does somewhat increase the frequency of illegitimate
matings among sibs.

TABLE 61

Symbol.	Frequency.	Offspring.			Proportion Incompatible.	
		L	M	S	Parent-Offspring.	Sibs.
Long and Mid						
$ms/ms = Ms/ms$	$4\sqrt{3}-6$	$\frac{1}{2}$	$\frac{1}{2}$	...	$\frac{1}{2}$	$\frac{1}{2}$
$ms/ms = Ms/Ms$	$7-4\sqrt{3}$	...	1	...	$\frac{1}{2}$	1
Long and Short						
$ms/ms = mS/ms$	$4-2\sqrt{3}$	$\frac{1}{2}$	...	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$ms/ms = MS/ms$	$3\sqrt{3}-5$	$\frac{1}{2}q$	$\frac{1}{2}p$	$\frac{1}{2}$	$\frac{3}{8}$	$(1-pq)/2$
$ms/ms = Ms/mS$	$3\sqrt{3}-5$	$\frac{1}{2}p$	$\frac{1}{2}q$	$\frac{1}{2}$		
$ms/ms = MS/Ms$	$7-4\sqrt{3}$	...	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
Mid and Short						
$Ms/ms = mS/ms$	$28\sqrt{3}-48$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{8}$
$Ms/ms = MS/ms$	$66-38\sqrt{3}$	$\frac{1}{4}q$	$\frac{1}{4}(1+p)$	$\frac{1}{2}$	$\frac{7}{16}$	$(7-pq)/16$
$Ms/ms = mS/Ms$	$66-38\sqrt{3}$	$\frac{1}{4}p$	$\frac{1}{4}(1+q)$	$\frac{1}{2}$		
$Ms/ms = MS/Ms$	$52\sqrt{3}-90$	...	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$Ms/Ms = mS/ms$	$52-30\sqrt{3}$	...	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$
$Ms/Ms = MS/ms$	$41\sqrt{3}-71$	...	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$
$Ms/Ms = mS/Ms$	$41\sqrt{3}-71$	...	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$
$Ms/Ms = MS/Ms$	$97-56\sqrt{3}$	...	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

If for all fourteen types of mating the proportion of illegitimate parent-offspring matings is multiplied by the frequency and added, the general proportion of illegitimate matings is found to be

$$(6\sqrt{3}-5)/12.$$

The proportion legitimate is therefore

$$(17-6\sqrt{3})/12,$$

and, since with unrelated plants this proportion is two-thirds, the effect of the genetic system is to reduce the proportion of parent-offspring matings in the ratio

$$(17-6\sqrt{3})/8$$

or

$$82 \cdot 596 \text{ per cent.}$$

For sib-matings we have a corresponding reduction in the ratio

$$\{(1+3\sqrt{3})+(13-7\sqrt{3})pq\}/8$$

or, as a percentage,

$$77 \cdot 452 + (10 \cdot 946)pq.$$

For sib-matings, therefore, the proportionate reduction ranges from $77 \cdot 452$ per cent. for absolute linkage ($pq = 0$) to $80 \cdot 588$ for very loose linkage, or independence ($pq = \frac{1}{4}$). For *Oxalis valdiviensis* the recombination fraction has been found to be about 7 per cent., and pq must be about $\frac{1}{15}$.

Compared with a distylic species, therefore, the tristyly in *O. valdiviensis* avoids about one in six parent-offspring matings, and a little more than one in five sib-matings. The latter effect is slightly enhanced by the close linkage of the two genes.

A second case of tristyly is supplied by *Lythrum salicaria*, (8) in which the loci for Mid- and Short-style are unlinked. Inheritance is, however, tetrasomic, and double reduction has been shown to occur at both loci. At the locus for Mid the frequency of double reduction has been determined to be about 8 per cent., and, using this value, the frequencies of fifteen genotypes in a cross-pollinated population, with equal viability and

fertility, have been calculated to be as follows, for a population of 1000 plants.

TABLE 62

		Not Short, s_4.	Short, Ss_2 and S_2s_2.
Long	m_4	333·3333	179·6645
Mid	$M\,m_3$	275·9408	114·7611
Mid	M_2m_2	52·1439	33·6590
Mid	M_3m	4·9246	4·9246
Mid	M_4	0·3241	0·3241

Owing to double reduction at the Short locus a certain fraction of each class of Short genotype will be duplex for the Short gene. In equilibrium this fraction is

$$3a/(4-a),$$

where a is the frequency of double reduction at the Short locus. The frequency a has been very roughly determined at about $2\frac{1}{2}$ per cent., so that about one in fifty-three of each class of Short will be duplex. This proportion does not affect the frequency of legitimate matings of parent and offspring, but would have some influence on the frequency for whole sibs.

The frequency of legitimate matings for parent and offspring may be expeditiously calculated from the frequency of the different style types from Long, Mid or Short seed-parents.

TABLE 63

Style Types of Legitimate Offspring from Different Seed-Parents

Seed-Parent.			Long.	Mid.	Short.
Long	.	.	429·96	370·04	250·00
Mid	.	.	322·47	427·53	250·00
Short	.	.	247·57	252·43	500·00

Whence it appears that the proportions of legitimate matings with offspring are :—

TABLE 64

Seed-Parent.						Per cent.
Long	.	.	.	.	.	57·004
Mid	.	.	.	.	.	57·247
Short	.	.	.	.	.	50·000

Thus, on the average, $54 \cdot 750$ of parent-offspring crosses are legitimate, or $81 \cdot 175$ per cent. of the proportion in the general population.

The protection afforded against matings of parent with offspring in *L. salicaria* is thus very nearly equal to, though a trifle greater than in *O. valdiviensis*.

The effect of a series of self-sterility allelomorphs is less determinate owing to its dependence on the frequencies with which these occur. If all are rare, as seems often to be the case, then almost all matings will be of the type

$$s_1 s_2 = s_3 s_4,$$

and all genotypes of offspring will be cross-fertile with both parents. Since there are four intra-sterile types of offspring appearing in equal numbers, compatible sib-matings will be reduced to 75 per cent., or rather more than in *O. valdiviensis*. This form of self-sterility is, therefore, an efficient one for annual plants, though not so efficient as tristyly for perennials.

There will, however, be a proportion of matings of the type

$$s_1 s_2 = s_1 s_3,$$

giving only two genotypes, and having therefore 50 per cent. protection against sib-matings. In these crosses, moreover, half the offspring will be incompatible with the pollen parent, so that there will be only 75 per cent. compatibility in parent-offspring crosses. The large number of self-sterility alleles usually found will, however, make such matings comparatively rare.

In effect, then, a series of self-sterility alleles provides the greatest protection against homozygosity for annual plants, in which sib-mating is of importance, while tristyly may be more effectual in perennials. Tristyly also is not upset if the powerful additional protection of polysomic inheritance is superimposed on the self-sterility mechanism, as in *Lythrum salicaria*.

REFERENCES

(7) R. A. FISHER and V. C. MARTIN (1948). Genetics of style-length in *Oxalis*. Nature, clxii. 533.

(8) R. A. FISHER and V. E. MARTIN (1947). Spontaneous occurrence in *Lythrum salicaria* of plants duplex for the short-style gene. clx. 541.

APPENDIX C

THE FUNCTION OF INBREEDING IN ANIMAL AND PLANT IMPROVEMENT

Great practical success has attended the production of improved varieties, most conspicuously in maize, by the following cycle of operations :—

(*a*) Choice of foundation stock.
(*b*) Inbreeding to produce a homozygous, or nearly homozygous, line.
(*c*) Crossing chosen lines ; the first generation from a pair of lines may be further crossed either to a third line, or to another first cross from different parent lines. Further crosses may be made using independent inbred material. The result of such crosses is used for production, and not as a self-perpetuating variety ; it is produced anew as required from the inbred lines maintained permanently as parent stock.

In interpreting the working of such a cycle of operations it is important to observe that, starting with a cross-bred population containing numerous genotypes with definite frequencies, the whole cycle will merely reproduce these same genotypes with the same pro- babilities provided that :—

(*a*) The foundation individuals are chosen at random.
(*b*) Inbreeding is carried on without selection.
(*c*) Inbred lines are crossed at random.

Any improvement actually effected must therefore be ascribed to selection at one or more of these three stages.

Most animal and plant material of value to mankind owes its value to the selection in past generations of the

visible good qualities of individuals. Carried out over long periods, and by innumerable individual breeders, each careful to preserve his best stock, this is undoubtedly an effective method of improvement; it is indeed the necessary foundation for all improvement. Since, however, domesticated varieties do not very rapidly change their average performance, the dramatic improvement effected by inbreeding cannot be ascribed to the single act of the selection of the individuals used to found the line. Its success may, of course, owe much to the choice of good stock from which these individuals are selected, but improvement above the level already attained by this stock is often far greater than careful selection within it, for a single generation, could possibly effect. We cannot ascribe the success of the programme merely to the fact that the foundation individuals are not chosen at random.

It would be quite impossible to carry out an inbreeding programme without selection, for the numerous recessive defects which inbreeding uncovers, will by lethality or severe debility exercise a selection quite out of the experimenter's control. There are, however, grave difficulties of a quantitative nature, in the way of accepting the view that the success of inbreeding programmes is due simply to the purification of the stock by the elimination of a fraction of the recessive heritable defects.

In a cross-bred population carrying numerous recessive factors, the effect of each of which, when homozygous, is to lower somewhat the physiological efficiency, the general performance measured by yield of seed will be to some extent, but not greatly, depressed by the chance occurrence of such homozygotes. Suppose such a recessive gene to have a frequency p; in random crossing the frequency of the homozygotes will be p^2. If these have their yield lowered by the fraction k, the

average loss due to this fraction will be kp^2, and the total loss of crop from this cause will be

$$S(kp^2)$$

the summation being taken over all such recessives.

If no conscious selection is exercised in such a self-perpetuating cross-bred population each frequency p will come to equilibrium with the mutation rate, by which the recessive defect is produced. If the chance of survival is equated to the yield, as is reasonable with grain crops, the rate of elimination of the genes is also measured by kp^2, so that the sum of these quantities for all factors must be equal to the total of the mutation rates by which they are supported.

Since the great majority of mutants are known to be deleterious, we might now equate the depression of yield in the cross-bred crop to the total mutation rate to which it is subject, with, however, a few reservations. (i) Recessives causing complete lethality will contribute to the total mutation rate, but can scarcely affect the performance of the crop as grown, for this will be sown with excess seed, and the lethals will not appear at harvest ; the same is true of severely handicapped types, which will have been crowded out, yielding their space to more efficient competitors. (ii) Improved crops may now be regarded as working near to a " ceiling," or physical limit of production. As such a ceiling is approached each factor in the genotype has less and less effect upon the yield ; it is probable therefore that many frequencies are not now in equilibrium with their mutation rates ; since mutation may not have had time to increase these frequencies up to the new level of equilibrium. For both these reasons it would appear that the total elimination of deleterious recessives would make less difference to the yield of cross-bred commercial crops than the total mutation rate would

suggest. Perhaps no more than a 1 per cent. improvement could be looked for from this cause. Differences of the order of 20 per cent. remain to be explained.

Factors in which selection favours the heterozygote over both homozygotes will establish a stable polymorphism in which a considerable fraction of the population will be below the optimum. Such factors, if frequent, might explain a great advantage in some first-cross hybrids, but scarcely in later crosses, unless a multiplicity of alleles, all deleterious when homozygous, were assumed.

It should be noted that the phase of inbreeding is not favourable to the exercise of deliberate selection, partly because such selection often can only be applied at the expense of retarding the inbreeding process, partly because the effects of factors on which selection might usefully be exercised is masked by the segregation of a mass of deleterious recessives of no consequence to the final product.

At the third stage at which selection is exercised the conditions seem to be more favourable. We are not now confined to the physiologically deleterious factors, which by selection have become recessive. The difference between two different crosses may depend on any of the factors available for evolutionary adaptation. It may be noted first that " hybrid corn " has been an immense success in that species, in which thousands of inbred lines have been produced. In species in which so far only a few inbred lines have become available, success has not been conspicuous.

When many homozygous lines are available the conditions for effective selection seem to be at their best. (i) We are selecting the actual genotype to be used for production, not merely an ancestor of it. (ii) Owing to the reliability of breeding performance achieved by inbreeding, lots of all sizes will be available for testing.

Many crosses get no further than a first inspection, but promising crosses may be tested in quantity with all the precision which modern experimental design makes possible. (iii) Any special advantage, limited perhaps by locality or by industrial use, remains a permanent property of the hybrid, which will reappear whenever it is made up. It can later be produced in the quantity appropriate to the special role it is to play. Every careful determination of quality is a permanent contribution to the optimal utilisation of the material.

The practical moral of these facts is obvious. As the basis of future livestock and plant improvement there is required not a single inbred line, or a few only, but a deliberately planned multiplicity. The price paid for reliability of breeding behaviour is the impoverishment of the genic content, due to the elimination of many genes. There need be no such impoverishment if many inbred lines are created simultaneously.

STATISTICAL METHODS
FOR RESEARCH WORKERS

BY

RONALD A. FISHER, Sc.D., F.R.S.

ARTHUR BALFOUR PROFESSOR, UNIVERSITY OF CAMBRIDGE

The pioneer text-book for exact tests of significance, and analysis of Variance; fully illustrating a variety of computational procedures not available elsewhere.

Contents : Introductory. Diagrams. Distributions. Tests of Goodness of Fit, Independence and Homogeneity ; with table of x^2. Tests of Significance of Means, Differences of Means, and Regression Coefficients. The Correlation Coefficient. Intra-class Correlations and the Analysis of Variance. Further Applications of the Analysis of Variance. The Principles of Statistical Estimation. Sources used from Data and Methods. Bibliography.

10th Edition. Medium 8vo. xvi+356 pp. 16s. net.

THE DESIGN OF EXPERIMENTS

BY

RONALD A. FISHER, Sc.D., F.R.S.

ARTHUR BALFOUR PROFESSOR, UNIVERSITY OF CAMBRIDGE

'The introductory chapter discusses the grounds upon which evidence is disputed, the mathematical attitude towards induction and reasons for the rejection of inverse probability. In chapter 2 certain principles of design are illustrated by means of a simple experiment to test the veracity of a lady's assertion that she can tell, by tasting a cup of tea, whether milk or tea was first poured into the cup. The 3rd chapter consists of a re-examination of one of Charles Darwin's experiments to determine whether the difference in origin of inbred or cross-bred maize plants influences their growth rate, particular attention being drawn to the necessity of a valid estimate of experimental error. The 4th and 5th chapters deal with the randomized block, and the latin and higher square designs, respectively. New and complex extension of these to the factorial design, to the art of confounding and partial confounding, and to the value of concomitant measurements in design, are treated in the next 4 chapters. The 10th chapter on the generalization of null hypotheses and the 11th (final) chapter on the measurement of amount of information complete the work. Except for the last chapter, the treatment is essentially non-mathematical.'—*Biological Abstracts.*

4th Revised Edition. Demy 8vo. xii+240 pp. 12s. 6d. net.

***

OLIVER AND BOYD LTD.
EDINBURGH AND LONDON

STATISTICAL TABLES

FOR BIOLOGICAL, AGRICULTURAL
AND MEDICAL RESEARCH

BY

RONALD A. FISHER, Sc.D., F.R.S.
ARTHUR BALFOUR PROFESSOR, UNIVERSITY OF CAMBRIDGE

AND

FRANK YATES, Sc.D.
**HEAD OF STATISTICAL DEPARTMENT, ROTHAMSTED
EXPERIMENTAL STATION**

Since the publication in 1925 of Fisher's "Statistical Methods for Research Workers," exact methods for the treatment of small samples, and the combination of experimental data, have come widely into use. In addition to the tables in that book, which have been, by permission, widely published in other statistical text-books, additional tables auxiliary to the treatment of special types of biological, medical and agricultural data, have since been computed, and are here brought together in a single volume.

Within its limited compass it is the aim of the authors to present a collection which shall be, by itself, sufficient to meet the greater part of the needs of the modern statistician.

3rd Revised Edition.   Demy 4to.   viii+112 pp.   16s. net.

OLIVER AND BOYD LTD.
EDINBURGH AND LONDON

STATISTICAL METHODS IN RESEARCH AND PRODUCTION

WITH SPECIAL REFERENCE TO THE CHEMICAL INDUSTRY

Edited by

OWEN L. DAVIES, M.Sc., Ph.D.

" The book presents methods for dealing with the long tables of results or yields so often met with in large-scale production, on the basis of which the most economical working conditions must be chosen. The section on Specification will make an immediate appeal to those who had to produce during the war years hundreds of batches of material to comply with the exacting demands of the Chemical Inspectorate. This problem of specification is dealt with from the point of view of both producer and consumer, and a case is made out for statistical control as a substitute for rigid sample rules. A special section is devoted to the evergreen question of sampling. All the methods described in the book are clearly illustrated by examples selected from industrial chemical practice, and this adds reality to the very clear exposition. The authors are well aware of the necessity for eliminating extraneous variables before estimating degrees of correlation, while the use of frequency data to establish possible causes for accidents or mechanical failures is interesting.

" Throughout the work every attempt is made to lighten the mathematical tasks involved, and a spirited effort is made to transform what is often an unwieldy science into an easily handled tool."—*Times Review of Industry*.

Second Edition, revised. **304 pp.** **28s. net.**

PUBLISHED FOR

IMPERIAL CHEMICAL INDUSTRIES LIMITED

BY

OLIVER AND BOYD LTD.
EDINBURGH AND LONDON

.

354